TEMPORAL PROBABILISTIC MODELS FOR DISEASE MANAGEMENT

WITH APPLICATIONS IN COPD CARE

MAARTEN VAN DER HEIJDEN

ISBN 978-94-6108-523-8

Temporal Probabilistic Models for Disease Management

With applications in COPD care

Proefschrift

Ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 4 december 2013
om 14:30 uur precies
door

Maarten van der Heijden

geboren 25 juli 1984 te Nijmegen.

Promotor:	Prof. dr. P.J.F. Lucas	
Copromotor:	Dr. T.J.R. Schermer	
Manuscriptcommissie:	Prof. dr. ir. M.J. Plasmeijer	
	Prof. dr. ir. H.J. Hermens	Universiteit Twente
	Prof. dr. ir. M.J. Druzdzel	University of Pittsburgh

CONTENTS

1

INTRODUCTION

We make decisions all the time, almost unconsciously, relying on past experience and present observations. However, our experience might be insufficient and our observations imperfect, which introduces uncertainty. For instance, we may be unaware of some factors that would influence our decision had we known them. Therefore to make the right decision we somehow need to take into account the uncertainty in the information we have.

This thesis is about knowledge representation under uncertainty, applied to problems in biomedicine. Throughout the years knowledge representation has been a topic of artificial intelligence research. Here we present our take on part of the problem, focussing our efforts on challenges in the context of clinical medicine. The question we will ask ourselves is how we can represent knowledge about processes that are uncertain – for all kinds of reasons which we will detail later – in such a way that we understand what happens and can perhaps predict what will occur next.

1.1 MEDICAL KNOWLEDGE AND DECISION SUPPORT

The problem of knowledge representation under uncertainty is a general one and interesting in its own right. We will place emphasis on applications in biomedicine, which share many properties with general knowledge representation tasks but may involve other, for example ethical, considerations. Artificial intelligence research has also in the past often focussed on the problems of knowledge representation, uncertainty and decision making within medicine. Arguably the most famous application is MYCIN, a system advising on treatment of infectious diseases [88]. Basically, decision support systems somehow contain a description of a particular domain and, given observations, can suggest a possible course of action by reasoning with the domain knowledge and the observations. The details of how this is achieved have differed considerably through the years however.

Knowledge representation techniques

Logic has been a tool to structure knowledge dating back to ancient times, with the European roots attributed to Aristotle. The formal language that it is now is a powerful tool for representation and reasoning. It is therefore not surprising that logic is still used as a representation language for decision support systems. However, in logic a proposition is either true or false, which is too crisp a distinction in the real world where we have to deal with gaps in our knowledge and other kinds of uncertainty. MYCIN uses a large set of rules, which is in essence a logic representation, with *certainty factors* to represent uncertainty [88]. Various alternatives to deal with uncertainty have been developed, but suffice it to say that *probability theory* is now mostly accepted as a sound way to represent uncertainty ([18] makes a case for probability theory compared to other representations like certainty factors).

Probability theory is a consistent manner to characterise random phenomena – although the colloquial meaning of the word 'random' is often associated with a uniform distribution implying equal probability for each outcome, we here mean random in the technical sense of a process involving chance. Probability theory was originally developed as a way to describe games of chance, with early work in the 16th and 17th century by Gerolamo Cardano, Pierre de Fermat, Blaise Pascal and Christiaan Huygens. However, it turned out that probability is useful in a much wider context than rolling dice, as it can be used in general to describe uncertainty. There are two main views on the interpretation of probability theory, called the *frequentist* and *Bayesian* view, which differ mainly in whether a probability should be seen as a relative frequency of an outcome in a repeatable experiment – rolling dice is a typical example – or as a subjective degree of believe in a given outcome – e.g. the probability that you will read all of this thesis. The work by Pierre-Simon Laplace at the start of the 19th century is considered important in the development of probability theory and the Bayesian perspective in particular. Since then many powerful statistical and machine learning methods have been developed based on both frequentist and Bayesian probability theory.

Models for decision support

Irrespective of whether we choose to use logic, statistics or some other approach, we are interested in *models* that somehow represent part of reality. Typically, a model tries to capture the important aspects of a problem, in order to obtain a better understanding of the underlying process, to reason about it and make predictions about future behaviour. Deciding which factors are important is not a trivial task, especially when not all factors are readily observable. An application domain for which this property is often true is clinical medicine. Many clinical problems exhibit the kind of complexity that asks for smart models to capture the relevant information. It is therefore not surprising that much research in statistics and in artificial intelligence has focussed on trying to solve diverse clinical tasks ranging from diagnosis to prognostics. Recently, technological advancement has lead to new initiatives extending medical informatics and artificial intelligence applications to web-based and mobile platforms. As a result new opportunities arise, but at the same time many of the challenges of knowledge representation remain.

It turns out that even with the well-developed theory of logic and of probability, building a successful decision support system is an endeavour in which knowledge representation is one of the difficult problems that has to be tackled. In general this requires insight in the domain and relevant data. Adherents to pure data driven approaches rely on what the data can tell them, mostly under the assumption that finding correlations is sufficient for all practical purposes. Insight in the domain is then gleaned from the data only. Indeed, statistics provides powerful tools and these will also be used in the work described here. Nonetheless it is arguable that a causal understanding of the domain as an expert might provide has merit beyond what can be said on the basis of correlation. When it comes to ascertaining causality, controlled experimentation is key, but often not every relevant factor can be controlled. In this context research that tries to capture causality in probabilistic models is of interest [76, 89]. The resulting models do not usually allow inference however, which makes them unsuitable for practical systems. In our opinion both data and (causal) expert knowledge tell part of the story, and insight in the domain is an important part of making sure a decision support system does the right thing; at the same time statistical information can provide quantified information and offers the opportunity for knowledge discovery by finding new relations in the data.

The characteristics of processes and data – obtained from these processes – that will receive particular attention here are *uncertainty* and *time*. Uncertainty arises in data as a result of limited observability, measurement errors, and missing values. At a more abstract level uncertainty may mean that the process is not fully understood. It may be that not all factors influencing a process are known, or it may be unclear how variables interact. To give a clinical example, the cause of a disease may be the interaction of a particular kind of bacteria with some physiological system, but is usually also influenced by environmental factors. Which factors play a role in the disease and how different factors interact will be a source of uncertainty in a description of the disease process. The second aspect, time, is important because we are often interested in making predictions, which means we have to understand the dynamics of a process, how future states depend on the present and possibly past states. Clinically, if we can predict disease progress, we can anticipate and adjust treatment.

Chronic disease management with probabilistic models

In this thesis we will be focussing on the issues just described in the context of chronic diseases, where especially time is a pertinent characteristic as chronic diseases are by definition persistent. The disease that will feature throughout is chronic obstructive pulmonary disease – which will often be abbreviated to COPD. This chronic lung disease provides a concrete example of the challenges that in general play a role in chronic disease management. One of the aspects of chronic disease management is patient monitoring and the interpretation of the data resulting from monitoring. As in the general case, uncertainty arises as a result of measurement errors, incomplete knowledge of the underlying disease process or the inability to measure certain aspects because of costs, both monetary as well as in terms of patient well-being or simply convenience. The time aspect is a direct consequence of monitoring patients for a prolonged period, resulting in longitudinal data, which impacts how we can and should interpret these data. The chronic nature of the disease naturally leads to such a situation.

To represent disease processes we use a class of probabilistic models called *Bayesian networks*, which in essence are factorised representations of joint probability distributions. Bayesian networks are a particular type of probabilistic graphical models, which represent relations between random variables in a graph. In the case of Bayesian networks this is a directed graph which has the interesting property of facilitating a causal interpretation, where a direct causal influence can be represented in the graph as an arc. Although it is not true in general that a Bayesian network has a causal interpretation, in a clinical context it is useful to work with causal models as it makes it easier to understand what the model means, without requiring a complete technical understanding. The causal interpretation also facilitates hand-crafting a model in cooperation with an expert, which can be augmented with data to estimate the parameters.

The progressive nature of chronic diseases and the temporal data that result from monitoring imply that it is necessary to take time into account in our models. Dynamic Bayesian networks are an extension of Bayesian networks for discrete time processes, where the graph indicates relations between variables either as atemporal using arcs as before or as temporal arcs to represent influences over time. This means that the evolution of a process can be modelled in detail, which in turn can be used to make predictions. In Chapter 2 we will consider these models in detail.

Alternative techniques exist that might have been used instead of (dynamic) Bayesian networks. For example continuous time Bayesian networks [72] could have been used to model the temporal process. However, discreet models are generally

easier to interpret. Interpretability also provides a motivation for the choice to start with a discreet hand-crafted model without explicit time information. Dynamic Bayesian networks then form a natural generalisation, as they are in some sense just a static model that is repeated over time, whereas continuous time requires a change in representation to a distribution that can model time, for example an exponential distribution. Another option would have been partially observable Markov decision processes. Our choice for Bayesian networks was motivated by the desire to first capture the nature of the disease process which does not yet involve decisions. It may however be useful to reconsider partially observable Markov decision processes to extend the disease model with decisions. We aimed to solve the problem of COPD support in a practical manner, trying to develop a solution that can be applied in the current health care environment. As a consequence decisions on what techniques were used have not only been made on theoretical grounds but have also been influenced by considerations of practicality, for example the availability of software implementations.

1.2 ON EHEALTH

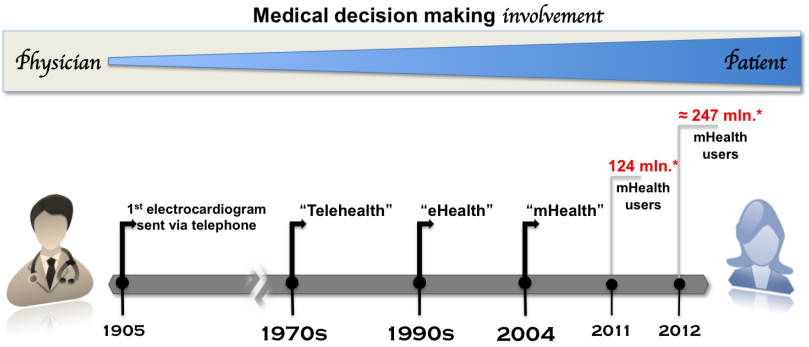
Computer assisted medicine has long been a topic of research in artificial intelligence and later in medical informatics. One aspect of this broad field of research is what is sometimes referred to as *telehealth*, or, *telemedicine*. One of the reasons for research in remote healthcare is that it is seen as a way of possible cost reduction, yet without sacrificing quality of care, which is one of the most significant current challenges of healthcare. Examples of measures that try to achieve these aims include reduction of the number of face-to-face consultations in out-patient clinics and prevention by detecting people at risk at an early stage and by providing guidance to prevent costly hospital admission. These last points have become a topic of much research under the denomination of *eHealth* (electronic health) and *mHealth* (mobile health). Modern names for variations on the same general concept of *facilitating health care irrespective of location by means of technology*. Concepts that are also often mentioned in this context are *patient empowerment* and *self-management*, as it is thought that eHealth solutions will enable patients to be more involved with and responsible for their own health.

A very short history

Figure 1.1 offers a visual summary of the development of the field. In general we see greater patient involvement combined with new technological advances.

A surprising, very early example of *telehealth* is the work by the Dutch physiologist and Nobel prize winner Willem Einthoven, the developer of the electrocardiograph. In 1905 he transmitted, via telephone wire, electric cardiac signals of a patient from the hospital to his laboratory 1.5 km away to demonstrate the power of his ideas. The next phase in telehealth was due to the development of the television in the 1950s, which allowed a patient's condition to be visualised at a distance. *Telemedicine* is a term introduced in the 1970s to indicate the distance management of a patient's condition, and thus includes a treatment component.

From the mid-1990s on, when internet access and usage became widespread and turned into the primary method for telecommunication, *eHealth* emerged as a promising field for better and more efficient healthcare delivery using web-enabled services. The internet-based foundation of eHealth has been reflected in the majority of 51 definitions of eHealth, reviewed in [73]. In comparison to telehealth, eHealth is a broader term and encompasses not only health services but also health information, education and research. Evidence for the subsequent rapid growth of eHealth sys-



Telehealth	Healthcare service delivery where physicians examine patients at a distance where information (e.g., voice, an image, a medical record, commands to a surgical robot) is exchanged using electronic means of communication
***e*Health**	Health services and information delivered or enhanced through the *Internet* and related technologies.
***m*Health**	The practice of medicine and public health supported by *mobile devices*

Figure 1.1: Timeline development and general definitions of telemedicine, eHealth and mHealth.

tems and tools is the development of a large body of systems to monitor patients' condition, usually in the home environment, and enabling earlier diagnosis and more effective treatment.

Many of the eHealth systems that are currently in use for patient monitoring are limited to transmitting data to a physician or nurse for manual interpretation (see e.g. [79] for a review of COPD telehealth). Alternatively, simple threshold detection is used to generate alerts. A more comprehensive view on different patient characteristics is not yet common. It therefore appears worthwhile to study automatic interpretation of monitoring data, taking multiple clinical factors into account, represented in a disease model.

Now, before going into the challenges of chronic disease management and monitoring in particular, we first introduce the chronic disease that will be the main focus of the clinical part of our research.

1.3 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Throughout this thesis we will consider specific examples of patient monitoring and data interpretation in the context of chronic obstructive pulmonary disease. COPD is a highly prevalent lung disease, that causes morbidity and mortality at significant rates worldwide. According to the World Health Organisation [110], currently around 64 million people suffer from COPD. Prospects are that this number is yet to rise, which makes it an important target for health care research.

COPD is caused mainly by inspiration of (tobacco) smoke, both actively or passively, and by exposure to severe air pollution. It is characterised by a combination of chronic bronchitis – airway inflammation – and emphysema – destructive enlargement of airspaces – leading to airflow limitation that is not fully reversible. It is a generally progressive disease that is currently not curable, but largely preventable and with proper treatment, manageable [104]. The main symptoms are a result of the chronic inflammation and airway obstruction and include dyspnea – difficulty

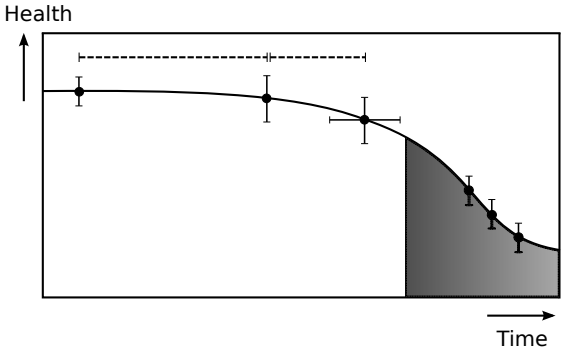


Figure 1.2: Abstract representation of patients monitoring data characteristics.

with breathing – sputum producing cough and wheezing breath. Because of reduced lung capacity patients also often suffer from fatigue and are sometimes unable to perform day-to-day activities due to shortness of breath. Clinical signs include reduced lung volume – usually measured by the forced expiratory volume in one second (FEV_1) – blood oxygen saturation level and body temperature – as indicator of airway infection induced fever.

The chronic and progressive nature of the disease make a case for monitoring patients. This is especially relevant for the group of patients that have frequent *exacerbations* – acute events of worsening of symptoms. Exacerbations play an important role in the disease progression, have significant impact on patient well-being and determine for a large part COPD related health-care costs [108]. Infection with respiratory viruses and airway bacteria is one of the main underlying causes of inflammatory reactions in the lungs resulting in the increase in symptoms of an exacerbation. Monitoring for early detection of exacerbations is useful as prompt treatment appears to lead to improved recovery and better health status [109]. In addition, if hospital admission can be prevented this benefits patients and reduces health-care cost.

1.4 PATIENT MONITORING DATA AND REASONING

The following situation sketch gives an idea of the general clinical case of chronic disease care and the challenges that arise, both theoretically and practically. In Figure 1.2 a graph is shown with time on the horizontal axis and on the vertical axis an abstract measure of health. This is a typical situation in chronic disease management, where one tries to maintain patient health and react appropriately to deterioration of health. The grey shaded area on the right indicates an episode of health status decrease that requires intervention. The nature of monitoring data results in a number of challenges.

- Measurement uncertainty
  Data may contain measurement errors, or uncertainty may arise from lack of knowledge about the exact relation between observable properties and the underlying causal mechanism (depicted as vertical error bars on the data points).

- Data sparsity
  It is often difficult to obtain sufficient data due to costs or inconvenience to

the patient. Additionally, data may not be available at crucial moments (note the lack of data points at the start of the shaded area).

- Irregular data acquisition
  Taking measurements may not always be possible at regular intervals, which may introduce time related confounding factors and complicates modelling the process through time (illustrated by the dashed lines indicating time between measurements).

- Temporal indeterminacy
  The exact time of the measurement may be unknown. This can be caused by a delay between measurement and registration or by failing to remember when exactly something occurred, when asked about a past event (depicted by the horizontal error bar).

- Heterogeneity
  Clinical data are often of a diverse character. Some of the values will be discreet, others continuous. Ranging from subjective symptom data to lab results, they are not easily comparable. A related problem is that some of the relevant variables may change on very different time scales.

To properly take these properties of the data into account, we need different modelling techniques. Some of these aspects will be studied in more detail and possible solutions will be developed.

1.5 OUTLINE OF THIS THESIS

In this thesis we report on the results of the Aerial project, in which a system has been developed for COPD-exacerbation management. In addition more theoretical work on representing uncertain temporal processes is presented. We now give an overview of the chapters in this thesis and summarise their contributions.

Chapter 2: Preliminaries

In the next chapter a brief overview is given of some necessary technical preliminaries on probability theory and probabilistic graphical models.

Chapter 3: An autonomous mobile system for the management of COPD

In this chapter the disease management system for COPD is described in detail. It is shown how mHealth can be used to help COPD patients gain insight in their disease through the use of disease models for automatic data interpretation. The main contributions of this chapter are the identification of system requirements for monitoring COPD patients at home, and the development of a prototype system that satisfies these constraints. Our objective was to develop an mHealth solution that enables self-management by means of decision support, requiring a means of data acquisition that is sufficiently reliable and usable in an uncontrolled environment; and data interpretation that can handle uncertain and sometimes missing time series information. The system implements data acquisition through a smartphone application to gather subjective information and sensors for objective information. Automatic data interpretation is performed by a Bayesian network that allows estimating the probability of a COPD-exacerbation given the observations. The model was constructed through knowledge elicitation from a pulmonologist. This model does not yet take the temporal aspect into account, however this is addressed in

Chapter 5. To our knowledge, a mobile system capable of automatic data interpretation is a novel contribution to chronic disease management and COPD care in particular. This chapter is based on [97]. My contribution consisted of the data analysis and writing the paper.

Chapter 4: Automatic COPD-exacerbation prediction: findings from a home-monitoring pilot study

In Chapter 4 our monitoring system is studied in a pilot study with COPD patients from the target population of patients with frequent exacerbations. We showed that the system is usable for COPD patients and that it records the required information to assess patient health status. Based on the data that was gathered, the exacerbation prediction model turned out to predict, with high accuracy, gradual onset exacerbations and detects rapid onset exacerbations at an early stage. This chapter is based on [100]. My contribution includes a large part of running the pilot studies and writing the paper.

Chapter 5: Learning Bayesian networks for clinical time series analysis

In this chapter an algorithm to learn a Bayesian network from a limited amount of time series data that contain missing values is described. The relatively low frequency of exacerbation events is the main cause of the limited availability of data. The algorithm is applied to data gathered in the pilot study described in Chapter 4. The contribution of Chapter 5 is a detailed study of temporal Bayesian networks for exacerbation prediction, in comparison to static models. The result of structure learning is compared to a temporal variant of naive Bayes and to the expert-opinion based models constructed earlier. Also studied are techniques to overcome the problems due to the limited availability of data. This lead to dynamic Bayesian networks learned via structural expectation maximisation and model averaged variants over bootstrapped time series. Using dynamic Bayesian networks learned from sparse data for clinical monitoring-data interpretation is a novel approach that appears to work well according to a validation performed on an independent data set. This chapter is based on [101]. My contribution was the data analysis and the main part of writing the paper.

Chapter 6: Probabilistic reasoning with temporal indeterminacy

In this chapter it is studied how temporal indeterminacy can be represented within the framework of Bayesian networks. Different assumptions on properties of indeterminacy are explored and an example in COPD monitoring is given. Although a related problem of the discrepancy between measurement time and registration has been studied for temporal databases, we here focused on modelling the temporal uncertainty probabilistically. We showed that Bayesian networks can be used for a representation at two temporal granularities, abstracting observations at a precise granularity to a coarser granularity; and modelling different temporal patterns going from a coarse granularity to a more precise granularity. This chapter is based on [99]. My contribution consisted of theoretical development and writing the paper.

Chapter 7: A probabilistic logic of qualitative time

The main contribution of Chapter 7 is a general logic-based language capable of expressing structural relations between events, probabilistic relations that capture uncertainty and temporal aspects through the use of qualitative relations based on

Allen's algebra [4]. This probabilistic logic of qualitative time allows us to reason about processes with uncertainty that evolve over time, of which the temporal information can only be specified partially. Clinically, this situation occurs frequently, since the complexities of clinical situations often result in partial observability which also manifests itself in not knowing the precise timing of events. We showed that we can describe the COPD monitoring process within our language and that it generalises other frameworks like Temporal Nodes Bayesian Networks. This chapter is based on [98]. My contribution includes theoretical development and writing the paper.

Chapter 8: Discussion and conclusion

Finally, we discuss the main findings of this thesis in a more general context and conclude with an outlook on the future.

2

PRELIMINARIES

In this chapter we give a brief overview of some of the theoretical background of the formalisms used in this thesis. In particular we describe the basics of probability theory and probabilistic graphical models. As we could easily fill a book trying to provide comprehensive preliminaries, we will assume some mathematical background and limit ourselves to introducing some important concepts. References to more comprehensive sources will be given in the text below.

2.1 PROBABILITY THEORY

Probabilities are a standard way to formally describe uncertainties, viewed either from a traditional statistical perspective as samples of some experiment that have particular outcomes with some relative frequency; or from the Bayesian viewpoint of degrees of belief of obtaining a particular outcome. Both interpretations have their use in certain situations and we will use them somewhat interchangeably throughout this thesis, although if pressed to take a stance a preference for a Bayesian interpretation would surface.

We give a short overview of some formal concepts that will prove useful later on, for a complete treatment see e.g. [41].

Definition 2.1. *Let Ω be the sample space containing all possible outcomes of the experiment; $\mathcal{F}$, a set of subsets of Ω, the event space; and P a function $P : \mathcal{F} \to [0, 1]$, such that:*

- $P(\emptyset) = 0$;
- $P(\Omega) = 1$;
- *For any countably infinite selection of disjoint events $e_1, e_2, \ldots \in \mathcal{F}$*

$$P\left(\bigcup_{i=1}^{\infty} e_i\right) = \sum_{i=1}^{\infty} P(e_i),$$

then the tuple $\langle \Omega, \mathcal{F}, P \rangle$ is called a probability space.

The last property is called *countable additivity* which simplifies to *finite additivity* for discrete spaces:

$$P\left(\bigcup_{i=1}^{n} e_i\right) = \sum_{i=1}^{n} P(e_i).$$

Usually it is more convenient to work with random variables than it is to work directly with probability spaces.

Definition 2.2. *A* random variable *is a function* $X : \Omega \to \mathbb{R}$*, such that* $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ *for each* $x \in \mathbb{R}$*. The* probability distribution *of a random variable is the function* $F : \mathbb{R} \to [0, 1]$ *given by* $F(x) = P(X \leq x)$*. The following distinction can be made:*

DISCRETE RANDOM VARIABLE *A random variable which takes on a finite number of values, for which the* probability mass function *is defined as the function* $f : \mathbb{R} \to [0, 1]$ *given by* $f(x) = P(X = x)$*;*

CONTINUOUS RANDOM VARIABLE *A random variable for which* $F(x) = \int_{-\infty}^{x} f(u)du$*, with* $f(u)$ *the* probability density function*.*

As we deal mainly with discrete variables in this thesis, the following definitions will be given for discrete variables and the similar definitions for continuous variables are omitted.

For multiple random variables, the distributions can be combined in a joint probability distribution.

Definition 2.3. *For a vector of random variables* $X = (X_1, X_2, \ldots, X_n)$ *the* joint probability function *is defined as* $f : \mathbb{R}^n \to [0, 1]$ *given by* $f(x) = P(X = x) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$*.*

The *marginal probability distribution* gives the probability of a single element from X disregarding the other variables; and is defined as

$$P(X_i = x_i) = \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} P(x_1, \ldots, x_i, \ldots, x_n). \tag{2.1}$$

The probability of an event given knowledge of a related event is called the *conditional probability*, which is defined in terms of joint and marginal probabilities for disjoint sets of variable X, Y:

$$P(X = x \mid Y = y) = P(x \mid y) = \frac{P(x, y)}{P(y)}, \tag{2.2}$$

for $P(y) > 0$. Often it is useful to write the joint probability in terms of conditional probabilities, which also shows the symmetry of x and y:

$$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x). \tag{2.3}$$

This decomposition of the joint probability can be performed recursively, which is known as the *chain rule of probability*:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{i+1}, \ldots, x_n). \tag{2.4}$$

By rearranging terms in Equation 2.3 we obtain *Bayes' rule*:

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}, \tag{2.5}$$

after the 18th century mathematician Thomas Bayes. Bayes' rule, also called Bayes' theorem, can be explained by taking x to be a hypothesis and y observed data pertaining to the hypothesis; the theorem states that the probability of the hypothesis given the data is equal to the *likelihood* of the hypothesis $P(y \mid x)$ times the *prior* belief in the hypothesis $P(x)$ normalised by the *evidence*, the probability of the data, $P(y)$.

2.2 PROBABILISTIC GRAPHICAL MODELS

Probabilistic graphical models graphically represent (conditional) independence relations between variables. Two sets of variables X, Y are said to be *conditionally independent* given Z, denoted $X \perp\!\!\!\perp_P Y \mid Z$, if

$$P(X \mid Y, Z) = P(X \mid Z).$$

Bayesian networks

In this thesis we will focus mainly on a specific class of models called *Bayesian networks* [75].

Definition 2.4. *A* Bayesian network *is a pair $\mathcal{BN} = \langle G, P \rangle$, with G a directed acyclic graph and P a probability distribution. The graph $G = \langle V, A \rangle$ consists of a set of vertices V, corresponding one-to-one with random variables X, and arcs $A \subseteq V \times V$ representing relations between the variables. Let $X \in V$ be a variable and $\mathrm{pa}(X)$ the parents of X in G. The distribution P is specified by a conditional probability distribution for each $X \in V$, of the type $P(X \mid \mathrm{pa}(X))$.*

A Bayesian network represents the joint distribution over a set of random variables X, which are represented by the vertices in the graph G. An important feature of the graphical representation of a Bayesian network is that the graph is an *independency map* (I-map) which means that the independences implied by the graph also hold in $P(X)$. That is,

$$X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp_P Y \mid Z,$$

where in a directed graph $X \perp\!\!\!\perp_G Y \mid Z$ can be established with the graphical criterion *d-separation* (see e.g. [75]). As the graph implies conditional independences, the joint probability can be factored according to the graph, as follows:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{pa}(X_i)), \tag{2.6}$$

where the $X_i \in X$ denote the random variables represented in the graph G.

Dynamic Bayesian networks

A *dynamic Bayesian network* [28, 69] is an extension of a Bayesian network to a distribution over a sequence of random variables, which is particularly well suited to representing time. A dynamic Bayesian network, DBN henceforth, represents a Markov process:

$$X_1 \to X_2 \to \cdots \to X_n \to \cdots$$

where X_t represents a random variable at a particular moment in time t. A hidden Markov model, an often used model for such a sequence, is a special case of a DBN. The joint distribution can be decomposed using the chain rule, writing $X_{1:T}$ for $X_1, \ldots, X_t, \ldots, X_T$:

$$P(X_{1:T}) = P(X_T \mid X_{1:T-1}) P(X_{1:T-1}).$$

In general a DBN is a factorisation of a probability distribution, like its atemporal counterpart. That is in the sequence above, each X_t can be represented as a BN that factorises the probability distribution over the values of the random variable X_t. This BN is called a *time slice* and relations between time slices can be modelled

by introducing arcs in the graph between random variables in different time slices. A DBN factorisation can then be written as:

$$P(X_{1:T}) = \prod_t \prod_i P(X_{t,i} \mid \text{pa}(X_{t,i}))$$

where i indexes variables within a time slice and $\text{pa}(X)$ denotes the parents of X in the graph.

A common assumption is that there is only a limited time frame that influences the current state of the process. Assuming a nth-order Markov process we obtain:

$$P(X_{1:T}) = P(X_T \mid X_{T-1:T-n}) P(X_{T-1:T-n})$$

When only limited data is available, it is common to make the most restrictive version of the Markov assumption, that is first-order, such that the future state of the process only depends on the present:

$$P(X_{t+1} \mid X_{1:t}) = P(X_{t+1} \mid X_t).$$

Hence all parents of a variable X will be in the same time slice or in the previous time slice. When we now also assume the process is stationary, that is $P(X_{t,i} \mid \text{pa}(X_{t,i})) = p$ for all t, we obtain a *two-slice DBN* consisting of an initial network BN_0 and a transition network $\text{BN}_\rightarrow$. A process through time can now be modelled by a sequence of repetitions of transition networks.

2.3 MODEL LEARNING

Bayesian networks can be constructed by hand on the basis of domain knowledge. Causal knowledge can be used to structure the graph while the parameters can be estimated from facts available in the literature or directly by a domain expert. This is however a time consuming process, and depending on the size of the domain may or may not be feasible. The model reported in Chapter 3 is constructed with expert knowledge. The alternative is a statistical approach to the problem: learning a model from data. There are multiple ways in which data can be used in model construction: parameter estimation in a model with predetermined graph structure; full model learning of both structure and parameters. Here we give a brief overview of some notions related to model learning, Chapter 5 provides a fairly detailed description in the context of Bayesian networks.

2.3.1 *Parameter learning*

The essence of learning parameters is finding the maximum likelihood estimate, or the Bayesian equivalent of finding the posterior distribution over the parameters. Given data D the maximum likelihood estimate of parameters θ is:

$$\hat{\theta} = \underset{\theta}{\text{argmax}}\, P(D; \theta). \tag{2.7}$$

The maximum likelihood parameters are those parameters that best describe the data. For example, if we take a binary random variable X and sample N identically and independently distributed data points, the maximum likelihood estimate of the parameter θ of the Bernoulli distribution $f(k; \theta) = \theta^k (1-\theta)^{1-k}$ is the expectation, $\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(X = 1)$.

Maximum likelihood estimates work well when large amounts of data are available. In practice we may not have sufficient data, in which case maximum likelihood can be unreliable. For example, the estimate of the bias of a coin that is in fact fair $\theta = \frac{1}{2}$

from three tosses that happen to land heads is $\theta = 1$. Although this can be alleviated with smoothing to obtain $\theta = \frac{3+1}{3+2} = 0.8$ where we used Laplace smoothing, also known as add-one smoothing, a Bayesian approach allows one to incorporate prior knowledge explicitly. In this case we could assume that in fact the coin is fair. The Bayesian parameter estimate is given by:

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \tag{2.8}$$

by Bayes' rule, where $P(D \mid \theta)$ is the likelihood and $P(\theta)$ the prior distribution over the parameters.

The Bayesian perspective has the advantage that we can start from a sensible distribution on the basis of expert knowledge or from a uniform prior when no specific information is available. As Equation (2.8) shows the result of Bayesian parameter estimation is a distribution over the parameters instead of a point estimate, essentially modelling the uncertainty in the estimate. Sometimes a point estimate is also used in the Bayesian case; the *maximum a posteriori* estimate is $\max P(\theta \mid D)$, which is different from the maximum likelihood estimate as it is weighted by the prior distribution.

One of the problems with learning parameters from data is missing data values, which often occur in practice. If values are missing for reasons independent from what we are measuring this is called *missing completely at random* (MCAR). In this case the estimate is unbiased if we compute it from the available data. Unfortunately missing values are seldom MCAR, instead the missingness is correlated with the values of other variables, which is called *missing at random* (MAR). If it depends on the value of the variable itself whether it is missing this is called *not missing at random* (NMAR), which introduces a bias in the estimate. Most learning algorithms assume data is missing at random, which is not always warranted, but works fairly well in practice.

The process of filling in missing values is called imputation. A simple approach is replacing a missing value with its mean, which is unbiased for a single variable. In general for multivariate estimation the correlations of the missing values with other variables should be taken into account to obtain reliable estimates. One approach is multiple imputation, where different data sets are generated with some (nondeterministic) imputation method and the final result is averaged over the data sets. The expectation maximisation algorithm [31] is an iterative procedure to estimate parameters from data with MAR-values. In the expectation step the current estimate of the parameters is used to compute the posterior probability over the missing values. The probabilities are interpreted as fractional counts when computing the sufficient statistics and subsequently the maximisation step finds new parameters based on the sufficient statistics. This is repeated until the estimates converge. See also Chapter 5.

2.3.2 *Structure learning*

A Bayesian network structure implies certain dependences and independences between variables. There are two main approaches to the problem of learning network structures: constraint based methods and score based methods. Constraint based methods try to directly test whether variables are independent with a series of statistical independence tests. This works well if the data satisfies the criteria of the test. Especially for small data samples criteria like normality may not hold in which case score based methods might be preferable. Structure scoring relies on some measure over network structures and a Bayesian approach then leads to the probabilistic score

$$P(G \mid D) \propto P(D \mid G)P(G),$$

which is the posterior probability of the graph structure G given the data D. Computing the likelihood term $P(D \mid G)$ requires learning the parameters. As the number of possible graph structures grows exponentially with the number of variables, it is for most realistic domains not possible to search exhaustively. Instead a greedy search is used that adds, removes or reverses the direction of an arc in each step, choosing the operation that improves the score the most.

Greedy search procedures have the disadvantage that finding the optimal solution is not guaranteed, although restarting the search from different initial configurations helps cover the search space. In general it depends on the kind of problem and amount of data whether the optimal network can be found. Structure learning from data with missing values is a particularly difficult task because to estimate the missing values correlations have to be taken into account. In Chapter 5 we address some of these issues. A detailed account of learning Bayesian networks from data is given in [57].

3

AN AUTONOMOUS MOBILE SYSTEM FOR THE MANAGEMENT OF COPD

3.1 INTRODUCTION

Increasing demands on health-care and continuous pressure from health-care authorities and insurance companies to reduce costs while maintaining quality of care has created a situation in which automation of particular parts of the patient's care process has attracted attention. Especially the provision of computer-aided assistance in the management of the patient's diseases is an attractive option. In the context of chronic diseases, patients are continuously at risk of deterioration of health, requiring regular medical checkups and monitoring of their health status by the treating medical doctor. Providing computer-aided support to the patient can relieve work-load of health-care workers, while helping patients self-manage their disease. However, the provision of computer-aided support to patients poses questions with respect to whether or not patients are able to profit from the support, how the patient data needed for that purpose can be collected and interpreted, and which technical infrastructure is most effective.

The idea to offer computer-aided support to patients at a distance from the treating hospital or practice is not new. Remote care facilitated by telecommunication technology has existed for some time already under different names such as: 'telehealth', 'telemedicine' and more recently 'eHealth' (electronic health) and 'mHealth' (mobile health). The exact scope and definition of all these terms varies, as exemplified by the review of Oh et al. [73] that identifies 51 definitions of the term 'eHealth'. Although different definitions will place different emphases we can summarise the definitions as *facilitating health care irrespective of location by means of technology*. Practically speaking this means that information and communication technology assists in disease management, patient-doctor communication, patient education or any other application that promotes health.

The research described in this paper aimed at developing methods for computer-aided assistance, including event detection, alerting, monitoring and treatment advice, as part of chronic disease management at a distance from the hospital. Patients with diseases such as diabetes mellitus, chronic obstructive pulmonary disease (COPD), asthma and heart failure can benefit from assistance. We have applied it here to assist COPD patients, but the general framework is applicable to many other health-care situations. For COPD we collect data on respiratory symptoms, measure lung function and interpret these data by a probabilistic model to assess the risk of a clinically relevant worsening of symptoms due to an exacerbation.

The main challenge addressed by the research was to develop a computer-aided disease management framework that allows finding a proper balance between self-management by the patient and support in various forms by health-care workers. This depends on the patient's demands and wishes, the disease being managed and the requirements of health-care workers. One of the implications of the resulting requirements was that disease-related patient data had to be interpreted automatic-

ally as part of the patient's self-management support. The system had to be capable of delivering autonomous assistance in disease management without being intrusive. Hence we decided for data monitoring and interpretation directly on a smartphone. Thus, the smartphone became the device of choice, enabling instant feedback to the patient. To cover the whole spectrum from complete self-management to distance support by health-care workers, collected patient data can also be relayed to health-care providers through a central server for both disease management configuration and inspection.

Use of a smartphone makes it possible to forego the need for a personal computer (PC) with internet connectivity. This has the advantage that whereas most people are used to responding to phone alerts, sending reminders via email or a website may have little effect on the patient's behaviour. This is also very much in the spirit of health care that is no longer tied to a particular location.

If we compare this to usual care we see that patients have regular but fairly sporadic contact with their physician unless there is an acute reason for an unscheduled visit. Our system is capable of advising patients to take measures at an earlier stage than usual and enables easier communication between patients and health-care providers. Although parts of what the system provides has been used before in the context of eHealth or decision-support systems meant for health-care workers as user, the idea to place both healthcare users and workers on an equal footing by offering patients sophisticated, mobile decision support is new. The more common eHealth systems are mainly focussed on sending small sets of measurements from the patient to health-care providers, who still have to interpret these findings [67]. In the system described below, clinical knowledge is incorporated by means of a probabilistic graphical model in the smartphone, making it possible to provide relevant clinical advice automatically, to assist in patient self-management.

The following questions are explored in the rest of the paper. It is investigated in what way patients can be empowered with disease management assistance to prevent exacerbations of COPD. Another relevant question is whether it is feasible to automatically interpret monitoring data by probabilistic models to detect clinically relevant events. This question is explored by means of statistical model validation methods. As a prerequisite we examine choices with respect to the important features of the disease management system in terms of hardware and software. Finally, the usability of the system was investigated with the help of COPD patients. Summarising, we report on a system for COPD exacerbation management, that has the novel feature of including automatic data interpretation by a probabilistic risk model, enabling autonomous operation to support patient self-management.

In the next section, we start with some background on the clinical problem addressed in the research. Then the architecture of the disease management system is described in Section 3.3, focussing on its design and technical capabilities. Furthermore we report on a pilot study that investigates technical and clinical feasibility with a number of patients. The probabilistic model that is used for data interpretation is explained and evaluated in Section 3.5. Then in Section 3.6 we compare our system with telemonitoring requirements laid out by Peirce et al. [78] and with existing work on COPD telehealth, followed by a general discussion in Section 3.7.

3.2 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Chronic obstructive pulmonary disease, or COPD for short, is a chronic lung disease with high impact on patient well-being and with considerable health-care related costs [104]. Exacerbations – acute events of worsening of COPD-related health status – are important events in the progression of COPD, such that monitoring patients in a home setting to detect exacerbation onset may be warranted [66]. In this paper

we focus on disease management and specifically on detecting and managing the occurrence of exacerbations of COPD at an early stage. We aim to decrease the impact of COPD on the patient's quality of life, and prevent unscheduled doctor visits and hospitalisation due to exacerbations.

Chronic obstructive pulmonary disease is estimated to affect some 64 million people worldwide [110] and is one of the major chronic diseases in terms of both morbidity and mortality. COPD affects the respiratory system, decreasing lung capacity and obstructing airways, thus interfering with normal breathing. Patients often suffer from a combination of emphysema and chronic bronchitis, causing shortness of breath and therefore reducing their capability of performing day-to-day activities. The main cause of COPD is prolonged exposure to tobacco smoke; other causes include severe air pollution. COPD is currently not curable, but treatment does reduce the burden considerably. For further information on COPD see e.g. the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [40].

An important aspect of COPD which is particularly relevant in the present context is the progressive nature of the disease. Specifically exacerbations have a profound impact on the patient's well-being and on health-care costs [104]. These exacerbations are mainly caused by airway infections resulting in symptom worsening [108]. Important to note is also that patients with frequent exacerbations usually have faster disease progression, which makes exacerbation prevention a particularly relevant goal. Additionally, a faster treatment response to exacerbations leads to better recovery [109]. We can distinguish different clinical approaches to defining exacerbations. Due to limited observability it is not feasible to give a practical definition in terms of pathophysiology, hence exacerbations are usually defined as an increase in symptoms; in terms of use of medication; or in terms of unscheduled health care use. We will return to this point in Section 3.5.

The state of the respiratory system is observable via symptoms including dyspnea (breathlessness), productive cough, wheezing breath and decreased activity due to breathlessness. Besides these symptoms a number of physiological signs are relevant, in particular the forced expiratory volume in 1 second (FEV_1) and blood oxygen saturation. FEV_1 measures airway obstruction by testing to what extend the patient can overcome obstructive and restrictive resistance during forced exhalation. A number of other indicators of deterioration exist, like blood oxygen pressure, inflammatory proteins and white blood-cell counts. However, measuring these factors requires hospital-grade equipment and incurs considerable inconvenience for the patient. Blood oxygen pressure can be observed by proxy with a pulse-oximeter that measures blood oxygen saturation.

3.3 REMOTE DISEASE MANAGEMENT

The long term nature of COPD and associated exacerbation risk require that any system, deployed in a home-care setting, takes into account not only efficacy, but also usability as important factors in the design. This section describes the current system design, the choices we made and some of the issues that arose during the development and implementation.

3.3.1 *System overview*

GENERAL ARCHITECTURE In Figure 3.1 a graphical representation of the general idea behind our disease management setup is shown. The system consists of a smartphone as the main component taking care of communication and computation. Sensors are used to obtain objective information on the patient's health status, transmitted wirelessly to the smartphone. Questionnaire data is collected from the

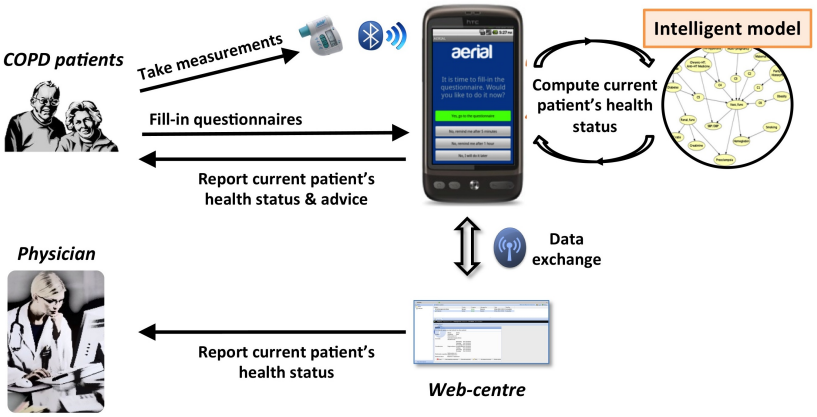


Figure 3.1: Schematic of the system setup.

patient on the smartphone. A web-based system allows scheduling tasks and collecting patient data centrally. The web-centre receives the data from the smartphone and provides data access for health-care workers. Patient data are interpreted in the smartphone by means of a disease-specific probabilistic model that incorporates clinical variables considered relevant by a clinical specialist.

THE COPD-SPECIFIC SYSTEM The system we used for COPD management is a specific instantiation of the general system just described. For COPD we were interested in measuring lung function and blood-oxygen levels to complement the recorded clinical symptoms. To do so we employed a sensor interface that took care of the Bluetooth communication and to which a micro-spirometer and pulse-oximeter were connected.

Before going into more detail on the various components let us first describe the monitoring process. At regular intervals, adjustable in frequency and in time of the day, the patient gets an automatic reminder for data entry from the smartphone. The patient is presented with a simple touch-interface to answer a set of questions about COPD symptoms and is subsequently asked to perform a spirometry test and pulse-oximeter measurement. The results of the measurements are transmitted to the phone, and entered in a Bayesian network model to determine the probability of an exacerbation. In addition, the data is synchronised with the web-centre, which allows the responsible health-care workers to examine the patient data; depending on the situation this may be a nurse specialised in lung diseases, general practitioner or pulmonologist. If necessary, the patient can be advised to take action, based on the model's prediction.

Monitoring patients for relapse or a new occurrence of exacerbation provides valuable additional information compared to usual care, as exacerbations are often underreported, with patients only contacting their physician when the exacerbation is already quite severe. For the population of patients with regular exacerbations, this intervention provides a means for the patients themselves and for health-care professionals to stay better informed on the patient's health status and to act accordingly.

3.3.2 *Design considerations*

We now discuss a number of design issues that are relevant for automatic disease management. Although we focus on the requirements for our COPD system, these points also have more general value for similar systems.

POPULATION Chronic disease management is by its very nature a long term effort, which makes it particularly important to find a careful balance between costs and benefits. The intrusiveness of monitoring systems and costs, both monetary and in terms of patient time investment, require a precise definition of the target population. We argue that those that stand to gain the most from our COPD system are the patients with moderate to severe COPD and frequent exacerbations (more than 2 per year). These patients suffer greatly from the consequences of exacerbations, hence providing regular data to detect exacerbations in an early stage will in general be more acceptable. The most appropriate time to start the intervention would be directly after emergency treatment, because at that time the goal of preventing hospitalisation is clearly relevant for the patient. To increase the chance of success in general, psychological factors associated with intervention acceptance and technology acceptance in particular should be taken into account when introducing a disease management system.

SECURITY The system can operate autonomously in which case no data is transmitted. If data is communicated to health-care providers, all data transmission from the smartphone to the server is encrypted (HTTPS) to protect the privacy sensitive nature of the data. Also access-rights to the data in the web-centre are controlled and patients should give prior consent. Since these are general issues when working with patient data, we will not focus on them here, but they remain important.

USABILITY & INSTRUCTION Ease of use is a critical requirement for any system that has to be used on a regular basis for a prolonged period of time. Since the interval between exacerbations is usually in the order of months, one should take care to reduce patient effort to a minimum, lest patients would stop entering data due to it being inconvenient. The patient population is relatively old on average – possibly not very experienced with the kind of technology offered – hence to facilitate understanding the system provides the ability to do practice runs. Nurses will have a supportive role in training patients. Specifically, at the start of the monitoring process, the nurse will enter the relevant information in the web-centre and instruct the patient on the use of the smartphone and sensors. A practice run serves to check whether the process has been understood and also provides baseline measurement values. In addition to the verbal instructions, the patient receives a written step-by-step instruction for reference. Our initial tests showed that a written instruction may be insufficient and therefore visual instructions, for example with screen captures of the phone, should be provided as well. As nurses are often just as pressed for time as physicians and also cannot be expected to be experts on the technicalities of the system, we provided a support phone-number for both patients and the nurses to assist in resolving any problems they might have. When introducing technology, especially to older patients, instruction by example appears to be preferable to only written instructions.

DATA COLLECTION RATE Depending on the health status of the individual patient the rate of data acquisition can be varied, which can be automated based on the acquired data and the model. As long as a patient has low risk of an exacerbation, monitoring can take place on a weekly basis, keeping the time investment at a minimum. If a patient is at risk according to model predictions, the system check-in

can be scheduled daily to ensure the possible exacerbation is detected and acted upon appropriately. A week between registrations may however be too long as an infection can develop in a matter of days, which means that unscheduled registrations are also an important part of the monitoring process. To facilitate self-management, it is important that the patient is in control of the registration schedule and can initiate registrations in addition to the predefined schedule. In this way patients can check their probability of exacerbation (appropriately visualised) when they feel they might be at risk. Eventually, this would hopefully lead to patients who no longer need our supporting technology, instead being able to decide by themselves when to contact their physician. The system, thus, has a dual role, providing information to the patient and health-care workers and helping the patient assess their COPD related health status, for better control over their disease.

SENSOR DATA RELIABILITY One of the critical points in the current system is the lung function measurement, which provides objective data to augment the subjective symptom data. Whereas oximetry is a simple manoeuvre clipping the sensor on your finger and waiting for a few seconds, spirometry is harder to perform well. In a home-care setting there is a risk of obtaining little useful data. Although patients that would use our system are diagnosed with COPD and therefore have to perform regular spirometry tests, it remains of paramount importance to properly instruct the patient at the start of the intervention. To adhere to the spirometry standards the FEV_1 measurement is performed thrice and the best value is used. For the sake of usability we do not continue until three successful measurements have been made, as this would raise the bar unrealistically high. A reduction of the problem comes from the fact that we only take FEV_1 into account which is easier and far less strenuous for the patient than measuring forced vital capacity (FVC), as is customary in lung function testing, and requires complete exhalation. In general, the reliability of all data that is gathered should be considered. Sensors can provide objective data only when they are sufficiently reliable.

FEEDBACK A further consideration is the kind of advice that the system should provide based on the model prediction of exacerbation risk. Typical advice would be to contact the physician, but a useful next step would be to advise, for example, to increase the dosage of bronchodilator drugs, corticosteroids or to start with antibiotics. The exact implementation of different kinds of advice should be decided in cooperation with clinicians, in accordance with clinical guidelines. We return to this point in Section 3.7.

3.3.3 *Hardware components of the disease management system*

Smartphone

Currently our system runs on an HTC Desire smartphone as an application in the Android OS. In principle any Android phone with Bluetooth capability should suffice, which makes the platform fairly general. If available the fast 3.5G HSDPA connection was used, but due to the use of asynchronous communication and the relatively limited amounts of data being transmitted, data can also be transmitted to the central server when only a slower 2G GPRS connection can be realised. Recall that the self-management part of system also functions without any server connectivity at all.

Figure 3.2: The monitoring kit consisting of a smartphone, sensor interface and sensors.

Sensor interface

The phone communicates with the sensors via a Mobi, a Bluetooth-capable multichannel sensor-interface, from Twente Medical Systems International. In our case a Nonin pulse-oximeter and a custom micro-spirometer were connected to the Mobi. An important advantage of using the Mobi sensor interface is the availability of the communication-protocol specification, enabling us to integrate the sensor readings seamlessly into the Android application. Most of the other micro-spirometers on the market do not allow this, which makes them unsuitable for easy deployment in a home setting. The monitoring kit is shown in Figure 3.2.

Pulse-oximeter

The pulse-oximeter used in this study was a Nonin Medical 8000AA, which is an industry standard pulse-oximeter. SpO_2 accuracy is 70-100% ± 2 digits.

Spirometer

We used custom-made pneumotachograph micro-spirometer prototypes by Twente Medical Systems International. These spirometers were newly developed to interact with the Mobi sensor interface and have the advantage of providing raw data such that analysing the spirometer readings is possible without requiring external software. This enables tight integration with our application, which is difficult or impossible with most commercial spirometers on the market.

Measuring raw spirometric data requires that we do some signal processing. The raw data signal is processed on the smartphone to obtain FEV_1 values, i.e. forced expiratory volume in 1 second. In Figure 3.3 a schematic spirometric flow-curve is shown. The shaded area is the FEV_1, which is computed from the raw data by first determining the peak of the curve, then finding the lowest point in the half second prior to the peak and integrating the flow during one second from there.

3.3.4 *The* Aerial *application*

As the project in which the intervention was developed was called *Aerial*, the smartphone application inherited the same name. Android applications are written in Dalvik, a programming language that uses Java syntax, but is compiled to run on the Android OS without the Java Virtual Machine. The application provides the following functionality: a timed alarm to signal the registration, with on-screen buttons to delay the start of the registration; a touch-screen interface for the questionnaire, which consists of 8 yes/no questions (see Table 3.1); on-screen instructions for

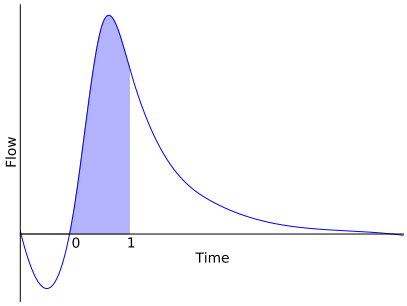


Figure 3.3: Schematic depiction of a spirometric flow-curve. Shaded area indicates FEV_1.

Since the last registration, did you have dyspnea worse than usual?
Since the last registration, did you have more sputum than usual?
Since the last registration, did you have sputum of a different colour than usual?
Since the last registration, did you cough more than usual?
Since the last registration, did you have more wheezing breath than usual?
Since the last registration, were you limited in performing day-to-day activities due to COPD?
Since the last registration, did you have a feeling of malaise?
Since the last registration, did you have a fever?

Table 3.1: Monitoring questionnaire. All on-screen text was in Dutch and has been translated to English for ease of exposition.

performing the measurements and Bluetooth communication with the Mobi sensor-interface to receive the measured data; processing of the spirometry data to compute the forced expiratory volume in 1 second (FEV_1); computation of the probability of an exacerbation based on the observed data (see also the section on the Bayesian network and model implementation below); asynchronously transmission of the observed data to the server over a secured data connection. In Figure 3.4 screen captures are shown of the alarm screen and of the question about cough.

3.3.5 *Risk model*

The main component of our system is the probabilistic risk model. Based on the data that is gathered, the model can compute the probability of an exacerbation. The model we use is a Bayesian network [75, 57]; Section 3.5 is devoted to the description of the development and evaluation of this model. The Bayesian network has been implemented using the lightweight reasoning engine EBayes [24]. Since EBayes is written in Java the model inference could easily be integrated in the Android application. Due to the relatively small size of the Bayesian network and the processing power of modern smartphones, the inference does not have to be deferred to a server but can be performed on site. This has the advantage that even when mobile phone network coverage is suboptimal the application can still provide a probability estimate and accompanying advice. Recall that this capability is also in line with our aim of patient empowerment as mentioned in the introduction.

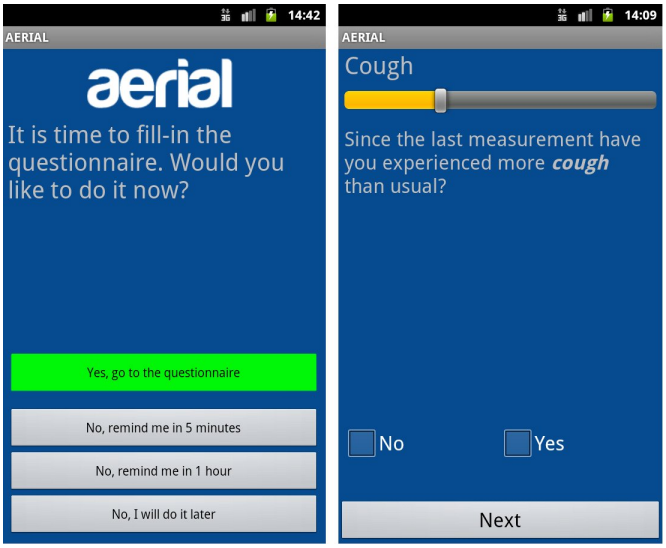


Figure 3.4: Screen captures of the application start screen at registration time and the question about cough.

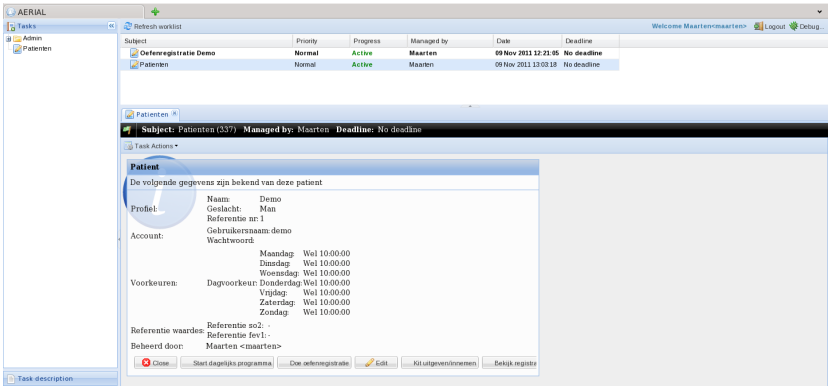


Figure 3.5: A screen capture of the patient administration web-interface.

3.3.6 *Web-centre*

The web-centre is the administration web-application that was built using the workflow management system iTasks [61]. The workflow system implements advanced features to generate and coordinate tasks and provides a generic (web)interface. Since the data management involved with patient monitoring is suitable to be represented as a workflow, iTasks provides a simple and effective way to construct the web-centre. In Figure 3.5 an impression of the web-centre is shown. Shielded by a secure login and different levels of access rights, the web-centre provides a workflow to enrol patients in the system as follows: first a user account is made with a registration number as identifier; the regular registration schedule is established in consultation with the patient; and a monitoring kit is assigned to the patient; then a practice registration can be performed (registrations are scheduled on the server; the smartphone regularly polls for new messages from the server but can also be forced to synchronise manually), of which the measurement results can be saved as baseline values; finally a start date and time for the registration schedule can be chosen. In addition, the web-centre has basic administration functionality to manage users and monitoring kit assignment as needed in research projects. Importantly, all the data received from the phone can be made available for inspection by the relevant health-care workers, thereby offering a means to stay apprised of the patient's health staus that is not available in usual care.

3.4 TECHNICAL FEASIBILITY AND USABILITY PILOT STUDY

Normally, there is a long way to go between the development of a prototype system and its actual use within a health-care setting. The involvement of patients in the development of a system is of crucial importance to improve a system based on their feedback and to ensure that the system meets the requirements.

We performed a system evaluation with a limited number of stable COPD patients recruited from the University Centre for Chronic Disease Dekkerswald in Groesbeek, the Netherlands, mainly to ensure that they were able to use the system effectively. We tested the monitoring equipment and data entry procedures with help of the patients and also studied the usability of the system from a patient's point of view.

3.4.1 *Methods*

For this study 5 stable patients were recruited by a convenience sample from the lung rehabilitation program who gave informed consent to participate in the study. Using three exacerbation monitoring kits (as described in Section 3.3) patients were monitored for a duration of 9 days, starting in January 2011. Patients were contacted daily during this time to answer a set of questions and perform spirometry and oximetry measurement. The answers were entered as evidence into the Bayesian network model to determine the probability of occurrence of an exacerbation. These predictions have not been used for patient advice yet, as further model validation is required. Patients were asked to report malfunctioning which together with the received data and server-logs could be used to verify system performance. At the end of the monitoring period a semi-structured evaluation interview was held – both with the patient and with the health-care staff involved – to obtain qualitative feedback on the usability of the system. In the interview we established whether or not the patients understood the procedure and the questions and whether they found the phone-interface sufficiently usable. Also, we checked anomalous data we received (if any) and asked for suggestions for improvement. The evaluation results of the first 2 patients were used to improve the system before starting with the second group.

As this was a technical feasibility test, the clinical data obtained were only used to check for errors in the application or obvious model inaccuracies.

3.4.2 *Results*

The pilot study allowed us to identify some problems in the server-side software, for example with respect to adequately recovering from connection errors, which could be amended relatively easily. After these amendments, the system functioned adequately at the technical level. With respect to usability, some problems became apparent at an early stage, such as the response buttons being placed too close together for comfortable use and an unclearly worded question resulting in confusion. In the evaluation interviews that were carried out with the patients the consensus was that the system could be useful to gain insight in the disease, was easy to use and not found to be intrusive. In particular patients indicated that they would be willing to use such a system in a home-care setting, which will need to be verified more rigorously in the next research stage. Thus, the patients' impression of the system after using it was quite positive and made us confident that the current system offers a suitable basis for moving toward a commercial product.

3.5 A PROBABILISTIC MODEL FOR THE DETECTION OF COPD EXACERBATION

The conclusion reached above, that the system was suitably functioning from a technical point of view, leaves the question unanswered whether it is also sufficiently accurate for predicting exacerbations, the ultimate aim of the development of the system. This question is addressed below.

3.5.1 *Bayesian networks*

For the data interpretation to detect exacerbations we use a *Bayesian network* [75], i.e. a probabilistic graphical model represented as a pair $\mathrm{BN} = (G, P)$. Here, $G = (V, A)$ is a directed acyclic graph consisting of *vertices* V, corresponding one-to-one to random variables of interest, and *arcs* $A \subseteq V \times V$, representing dependencies between variables. Furthermore, P is the quantitative part, denoting a joint probability distribution, specified in terms of a family of conditional probability distributions of the form $P(V \mid \mathrm{pa}(V))$, that is the probability that V takes on a specific value given the values of its parent variables. The network represents the joint distribution over the random variables, which can be factored according to the independences represented in the graph, resulting in:

$$P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P(V_i \mid \mathrm{pa}(V_i)),$$

where $V_i \in V$ is the representation of a random variable in the graph G. Probabilities of interest can be computed from this joint probability. In this case we are mainly interested in the probability of an exacerbation given the evidence obtained from monitoring.

Although other probabilistic models, such as based on logistic regression or artificial neural networks, could have been used for detecting exacerbations in patients, Bayesian networks appeared to be the most suitable method for our research. The main reason is their inherent flexibility, as it is possible to develop Bayesian networks from expert knowledge only, whereas learning of parameters and structure from data is also an option [63]. This is a necessary feature because of the ironic situation that although COPD is a very common disorder, there is currently not much patient data

available from which probabilistic models can be learned. This is caused by the early stage in which much of the eHealth research still is; few researchers have tried to monitor sufficient number of parameters of a sufficient number of patients at home in a sufficiently frequent way, e.g. more than once every week.

Bayesian networks also support reasoning about individual patients: although the model describes general relations between the variables of interest, all predictions are personalised by entering patient-specific data. The model is thus capable of making predictions for individual situations, and can provide 'what-if'-predictions by entering virtual evidence. Furthermore Bayesian networks are able to provide probability estimates given partial patient data.

3.5.2 *A Bayesian network for COPD*

For our COPD models, the main outcome variable is *exacerbation*, focussing on a symptom based definition. But the nature of a Bayesian network allows us to easily inspect probabilities for any variable. The network contains two hidden variables, namely *infection* and *lung function* which cannot be observed directly, but whose values can be derived based on indirect measures, such as body temperature for infection and the forced expiratory volume in 1 second (FEV_1) for lung function. Other important variables are the symptoms that one might expect a patient to report, such as dyspnea (breathlessness), sputum volume and purulence, cough, wheeze, general malaise, fever, and whether performing daily activities is difficult due to COPD. Additionally, the clinical signs FEV_1 and blood oxygen saturation (SpO_2) are included. Other signs and lab results are not included as these are generally not available in a home-care setting. All variables except for FEV_1 are binary and either encode presence or absence, or normal or abnormal states, which should be interpreted in comparison to the individual patient's baseline values. As FEV_1 is an important, objective measure it has 5 states: 'Normal'; -10%; -20%; -30%; <-30%; this division is based on expert opinion but would perhaps be represented more naturally as a continuous variable (as volume is clearly continuous), which requires appropriate continuous data.

Expert knowledge model

The goal of using expert opinion to construct the model was to obtain an understandable clinically justifiable model, in the absence of sufficient data that would allow learning the structure from the data. The contrast with a purely data driven approach lies in the fact that we can ask the expert for causal influences, instead of correlations obtained from data via statistical methods. We can therefore argue for the validity of the structure of the Bayesian network in causal terms, of course, without supposing that a Bayesian network is a causal network. We were interested primarily in the probability of an exacerbation, which is a clinically relevant event that is mediated by a number of physiological processes and influenced by environmental factors. We cannot hope to take into account every relevant factor, but it is clear that the symptoms of an exacerbation are a consequence of reduced lung function. Unobserved processes that constitute either normal or abnormal lung functioning can be represented by hidden variables. The measurements SpO_2 and FEV_1 are objective, but not sufficient to fully characterise lung function. The symptoms *dyspnea* and *wheeze* are direct consequences of reduced lung function. Whether patients are capable of performing daily activities also depends on whether the lungs are capable of providing sufficient oxygen. Next in the causal view are possible causes of reduced lung function. The most prevalent cause of an exacerbation is a lower airway infection, which is in itself a combination of unobservable processes, modelled by the variable *infection*. The presence of an infection is then the cause of

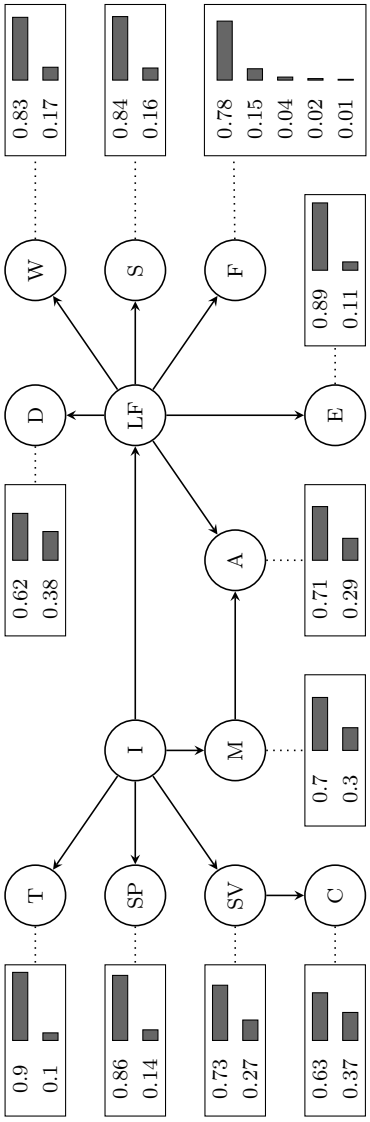


Figure 3.6: Expert opinion based Bayesian network with prior, marginal probabilities shown (top probability is Normal state). A = activity, C = cough, D = dyspnea, E = exacerbation, F = FEV_1, I = infection, LF = lung function, M = malaise, S = SpO_2, SP = sputum purulence, SV = sputum volume, T = temperature and W = Wheeze.

further symptoms like *fever*, general *malaise* and *sputum* production, the latter in turn causing cough to clear the sputum accumulation in the airways.

The resulting expert-opinion based COPD-exacerbation prediction model is depicted in Figure 3.6. The Bayesian network has been constructed in close cooperation with two pulmonologists (lung specialists) of the Radboud University Nijmegen Medical Centre. The relevant variables that are related to the occurence of exacerbations were identified by the domain experts and supported by recent clinical literature (e.g. [108]). As a second step the dependence relations between the variables were elicited from the domain experts, resulting in the qualitative part of the model. Subsequently, probabilities were estimated. As this is a fairly difficult task for clinicians not used to the abstract representation in terms of conditional probabilities, we started off by asking them to provide qualitative constraints of the type 'Is $P(X = \text{true} \mid \text{pa}(X) = x) <, =$ or $> P(X = \text{false} \mid \text{pa}(X) = x)$' (expressed in words). Based on these constraints an initial network was constructed to give some insight in how, approximately, the resulting Bayesian networks would look. We then went over different test cases to adjust the probabilities until the expert thought the resulting predictions were acceptable. The entire process has – in a different context – been previously described in [62].

Parameters from data

We have used two data sets to learn parameters and to validate the model. We started with a data set from an earlier project on detecting exacerbations at the Radboud University Medical Centre [12]. We will refer to this data set with A. This set consists of questionnaire answers of 86 Dutch COPD patients, 54 of whom had an exacerbation during the study. The data has been acquired biweekly during 2006-2007 via automatic telephone interviews, resulting in time series data with a total of 1922 data entries, of which 162 entries provided exacerbation data according to a symptom based definition of exacerbation. Later we will study the influence on the results by different definitions of an exacerbation. Data was only available for a subset of the variables in the model, specifically: exacerbation, dyspnea, sputum volume and purulence, wheeze, cough, and activity.

As a comparison we also validated the predictions on an unrelated data set [51], that provides time series of COPD-exacerbation related variables. We refer to this data set with B. This data set is used solely for external validation, so no model parameters were learned from this set. As the data is used for testing our model retrospectively, the set of variables is incomplete. However, it should give some indication of the generalisability of the results. The data consists of time series from 13 patients of the London COPD cohort who had an exacerbation, with a total of 2849 data entries, of which 406 were during an exacerbation. In contrast to the Dutch data set, these data were collected on a daily basis. The data contains values for the variables dyspnea, sputum volume and purulence, wheeze, cough, temperature, and SpO_2.

Parameter estimation proceeded as follows. Since not all data sets include all variables of interest, parameters were learned for those variables for which data was available. The usual approach to learning parameters from data entails computing a maximum likelihood estimate of the parameters, let us call them θ, for the given data D:

$$\operatorname*{argmax}_{\theta} P(D \mid \theta),$$

or from a Bayesian perspective, the maximum a posteriori (MAP) estimate:

$$\operatorname*{argmax}_{\theta} P(\theta \mid D) = \operatorname*{argmax}_{\theta} P(D \mid \theta)P(\theta)/P(D)$$

by Bayes' rule. Note that the denominator $P(D)$ is constant when maximising the expression for the parameter θ and therefore irrelevant. For complete data this can be computed easily when assuming that the data is independent and identically distributed (i.i.d.-assumption), but when data is incomplete – as it almost always will be in a clinical context – this should somehow be taken into account. We used the expectation-maximisation algorithm, popularised by Dempster, Laird and Rubin [31], which iteratively adjusts the parameters by alternating between estimating values for the missing data given the current parameters and using the completed dataset to compute new maximum likelihood parameters. Although this does not guarantee that the optimal parameters are found, Dempster et al. proved that the procedure improves the likelihood at each iteration, resulting in a local maximum of the likelihood function.

Comparison between expert and data models

We developed exacerbation predictions models both on the basis of expert opinion and from data. One would hope and expect that these models at least have some similarity. Although the parameters will usually not be the same we can make a qualitative comparison using the constraints we elicited from the expert. As described above we asked the expert to order the values of the random variables based on which would be more likely given the values of the parent variables. We would expect that the parameters learned from data would follow the same patterns.

For most of the variables for which we had data, the ordering constraints were met, except for the symptom variables *sputum purulence* and *activity* and the outcome variable *exacerbation*, which warranted more detailed inspection. For *sputum purulence* (SP) the qualitative constraints were that the conditional probabilities given *infection* were ordered as follows: $P(\text{SP} = \text{normal} \mid \text{I} = \text{false}) > P(\text{SP} = \text{abnormal} \mid \text{I} = \text{false})$ and $P(\text{SP} = \text{abnormal} \mid \text{I} = \text{true}) > P(\text{SP} = \text{normal} \mid \text{I} = \text{true})$. Analogous constraints were established for *activity* given *lung function*. When we look at the parameters learned from the data we can see that in both these cases the direction of the effect is correct and the probabilities are close to meeting the constraints. For instance the parameters for *sputum purulence* are $P(\text{SP} = \text{abnormal} \mid \text{I} = \text{false}) = 0.02$ and $P(\text{SP} = \text{abnormal} \mid \text{I} = \text{true}) = 0.43$, indicating that the probability increases considerably with the presence of an infection, but not quite enough to reach the limit of 50% which would reverse the ordering. The threshold of 50% follows from the qualitative constraints, because if for a binary variable X the probability $P(X = 0) < P(X = 1)$ then by definition $P(X = 1) > 0.5$. Something similar holds for *activity*. For the *exacerbation* variable the situation is a bit different in the sense that the distance to the threshold of 50% is larger, but the difference between the probability of an exacerbation given normal lung function or abnormal lung function is large. It appears that the qualitative constraint was too strict and there are more situations with decreased lung function that are not an exacerbation. However as the evaluation that follows will show, the difference between the probabilities for normal and abnormal lung function is sufficiently large to result in enough discriminative power.

3.5.3 *Model evaluation*

To gain some insight in model accuracy we can study the behaviour personalised with patient data. In the network of Figure 3.6 prior probabilities were shown, but by entering observations we can personalise the probability estimates to be consistent with a particular patient status. Table 3.2 contains the probabilities of some key variables, given different evidence situations. The causal nature of the model ensures

	Dyspnea	Sputum purulence	Cough	SpO_2	Exacerbation
No evidence	0.38	0.14	0.37	0.16	0.11
Cough	0.43	0.20	1	0.20	0.17
Cough, Wheeze	0.68	0.45	1	0.41	0.46
Dyspnea, SpO_2	1	0.51	0.53	1	0.55
Dyspnea, Sputum volume, FEV_1 (-20%)	1	0.67	0.80	0.60	0.71

Table 3.2: Model behaviour for different evidence situations. First column indicates which variables are observed to be abnormal. Other columns show probability of being abnormal (e.g. symptom is present).

that the probability of an exacerbation increases when more symptoms are reported or signs measured.

ROC-ANALYSIS To understand how the model behaves in general we performed an ROC-analysis, using 10-fold cross-validation with data set A, on the sub-model for which data was available to compare the expert opinion probabilities with learned parameters. That is, the variables in the set {exacerbation, dyspnea, sputum volume, sputum purulence, cough, wheeze, activity}. Prediction are obtained for each data point by inserting evidence into the network and inferring the resulting probability of an exacerbation. Concretely this first means that the observed variables are clamped to the observed value (and thus have probability 1). Second, the probability of an exacerbation is computed given the observations (using standard Bayesian network inference algorithms, see e.g. [75]). In Figure 3.8 the resulting ROC-curves are shown for both the data and expert model, conveying that the predictions can indeed distinguish the exacerbation cases. For the data model we find a mean area under the curve (AUC) of 0.93; and for expert model a mean AUC of 0.97.

Since the construction of our model we have obtained the data set (B) from Hurst et al. [51] described above. It is worthwhile to compare the results on this data set with the data set we used originally. To do so we looked at the subset of variables in B that intersect with the original data, that is {exacerbation,dyspnea,sputum volume,sputum purulence,cough,wheeze}. We then computed the model predictions for *exacerbation* given the observations on these variables and again performed an ROC-analysis. The area under the curve for these predictions was AUC = 0.87. As is common in ROC-analysis, we computed the optimal cut-off probability p^* by means of the point on the ROC curve closest to the $(0, 1)$ coordinates:

$$p^* = \underset{p}{\operatorname{argmin}}((1 - \mathrm{tpr}_p)^2 + \mathrm{fpr}_p^2)$$

where 'tpr_p' and 'fpr_p' are the true (false) positive rate that depend of the chosen threshold value p. The optimal cut-off value gives a general, problem-independent balance between true positive and false positive rates. Table 3.3 shows a comparison of the prediction results of the cross-validation model on the original and the new data in terms of AUC and the true (false) positive rate and accuracy at cut-off p^* (computed separately for each data set). As a further interesting evaluation we took the model learned from A and decided upon a cut-off probability which we then applied to the data from B. Hence we used the new data as if we did not have the correct labels, in the sense that we did not compute the optimal cut-off point. Since the A model was learned via cross-validation we first computed p^* for each of the cross-validation runs and compared two options, either taking the mean cut-off of the cross-validation runs p^A_{mean}; or the max cut-off probability p^A_{max} (row 5 and

Data set	AUC	TPR	FPR	ACC
A	0.93	1	0.16	0.85
A_{Drug}	0.81	0.79	0.3	0.72
$A_{Contact}$	0.85	0.87	0.3	0.72
B	0.87	0.88	0.20	0.81
B p^A_{mean}	-	1	1	0.14
B p^A_{max}	-	0.88	0.20	0.81

Table 3.3: Exacerbation detection performance for different data sets and definitions of exacerbation (see text for details).

6 in Table 3.3). Note that using the mean is too aggressive a cut-off as it results in all data being classified as an exacerbation. This can be explained by the class imbalance in the data, as misclassifying only a few data points has a relatively large influence on the performance measures. The optimal cut-off point is in some sense overfitted to the training data. Using the max cut-off however, we see that the performance is the same as for the optimal value for data set B. This sensitivity to the precise cut-off point will likely decrease when we obtain more data (especially positive examples).

DEFINITIONS OF EXACERBATION When using data to learn parameters it is important to note that clinically there are different definitions for when a worsening of COPD-related health status should be considered an exacerbation. The article by Bischoff et al. [12] that describes the study data we use here, also acknowledges this point by distinguishing exacerbations defined in terms of symptoms; drug prescriptions; and health care contact. In Figure 3.7 the counts are shown for data points according to the following definitions:

SYMPTOM-DEFINITION Presence for at least 2 days of at least the major symptoms dyspnea, sputum volume increase and sputum purulence.

DRUG-DEFINITION Start of a course of corticosteroids or antibiotics.

CONTACT-DEFINITION Unscheduled contact with health care workers due to worsening of COPD.

Interestingly enough there are many data points for which these definitions do not agree, which will of course influence model accuracy. Hence it is worthwhile to study the effect of using other definitions on predictions. Row 2 and 3 of Table 3.3 show the performance statistics of using the *drug* or *contact* definition with the model learned with the *symptom* definition. As expected this results in a considerable drop in performance, as the definitions capture different aspects of exacerbations.

More interesting is the performance of the expert opinion model as this is based on the experience and clinical gaze of the pulmonologist instead of a technical definition. The result, using the same indicators as before, is shown in Table 3.4. Interestingly enough we observe the same pattern here as with the learned parameters, in the sense that the symptom based definition performs considerably better than the event based definitions. A possible explanation is that drug use and health care contact data are based on self reporting which may have introduced noise (both over-reporting and under-reporting could have occurred due to for example misremembering). Since the symptom-based definition is derived from more variables it is less likely that someone

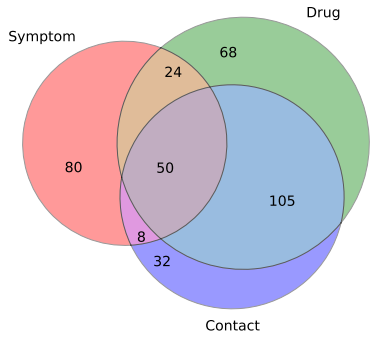


Figure 3.7: Venn diagram of data-point counts for different definitions of exacerbation in data set A.

Definition	AUC	TPR	FPR	ACC
Symptom	0.97	1	0.10	0.91
Drug	0.82	0.75	0.24	0.75
Contact	0.85	0.83	0.25	0.76

Table 3.4: Performance of the expert model for different definitions of exacerbation.

would mistakenly remember multiple symptoms. Another explanation might be that by explicitly constructing a model of symptoms we primed our expert to focus on the symptomatic part of exacerbations. Perhaps this balance would have been different had we included medication in the model. This is a topic for further investigation.

These results indicate that we can reliably detect exacerbations as they are happening, which is a practically useful result. The final trade-off between true and false positives should be based on more data, taking into account what definition of exacerbation is most useful for disease management. Compared to usual care it is an improvement even if we catch only some of the exacerbations that are now detected too late. Ultimately, the more difficult task of predicting exacerbations still lies ahead, as currently we lack sufficient data to model the temporal progression. In further testing of the system we also plan to gather the necessary data to construct a temporal model, which requires trend analysis of time courses of symptoms and signs leading up to an exacerbation.

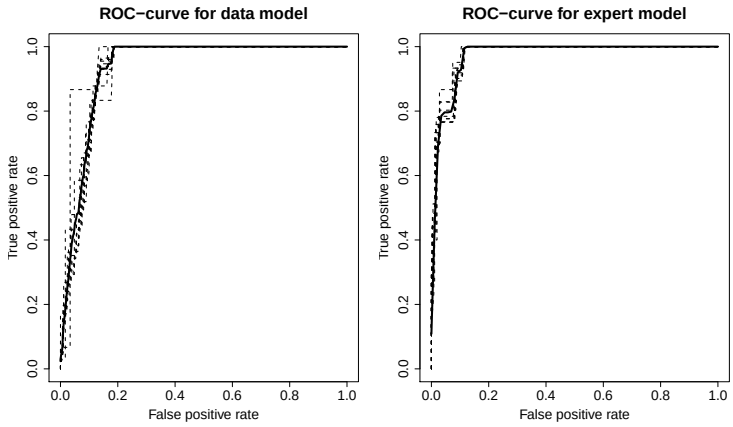


Figure 3.8: ROC-curves, cross-validation results (dashed lines) and average (bold).

3.6 RELATED WORK

3.6.1 *Telemonitoring requirements*

In a recent study by Peirce et al. [78] an extensive analysis of current telemonitoring practice was performed. Based on this analysis a set of requirement questions was identified that should be answered to obtain an effective telemonitoring system. Although we think that out system is more general than purely monitoring, it still appears useful to consider these questions and try to answer them for our system. *Why monitor?* Early detection of deterioration leads to improved outcomes for patients. *Who to monitor?* In general deciding on the target population is important for the success of an intervention. For our COPD system we target those patients that suffer from frequent exacerbations (more than 2 per year), which is a clearly defined group that stands to gain from timely intervention on exacerbations. *How to monitor?* Our system uses a combination of objective and subjective data, while trying to maintain a minimal effort requirement for patients. *What to monitor?* The variables in our model have been identified in cooperation with clinicians and the value of each variable is determined statistically from data to ensure we take into account the most relevant parameters. That is, we started out from clinical requirements, not from technological ones. However to obtain a usable system, in practice some concessions had to be made to what is possible technically and cost-wise. *How to detect deterioration early?* This is an important point that forms the basis of our current research. Now that we have a system that can detect exacerbations, the next step is modelling temporal progression in order to observe trends that allow us to predict an approaching exacerbation. Although Peirce et al. labelled early detection as a shortcoming of many systems, it is one of the main reasons for the development of our system. *How to display information appropriately?* We have tried to keep the user interface as simple as possible to accommodate our users who may not be proficient with modern smartphones. The presentation of the feedback based on model predictions is still a topic of further research. However as was recognised by Peirce et al. a combination of a simple feedback mechanism (say a smiley face, traffic light or some kind of simple colour coding scheme) with a more in depth analysis of the data (e.g. trend graphs for symptoms and measurements) will likely turn out useful. Patients can then choose the level of detail with which they are comfortable. *Who to*

notify? How to respond? Alert notifications are an important part of the practical embedding of a telemonitoring system in day-to-day care. We think that both the patient and their health care providers should stay part of the decision process, even if data interpretation is automated as it is here. Therefore, patients keep the initiative in contacting their physician, but the system can provide advice on when to do so. *How to store data?* It depends on the local health care structure whether integration with GP or hospital information systems is desirable. From the perspective of patient-empowerment/self-management, it is useful that our system can also be employed stand alone, leaving access rights with the patient. Summarising, our system is already well-equipped with respect to these telemonitoring requirements, yet refinements can also be identified by further attention to answering these questions.

3.6.2 *Related telemonitoring systems*

Monitoring COPD-patients has been a topic of interest for some time now, but the level of automation is usually limited. Basically we can distinguish two types of earlier work, telemonitoring where technology only provides a means of communication between patient and physician; and systems that have some kind of automated alerting or data interpretation. Systems of the first kind have for example been reported by [64], where pulse-oximetry data was send to the hospital via a telephone line and analysed by hand; [107], where a nurse set up a teleconference with a physician; and [106], where patients with respiratory failure were monitored by telephone. The system reported in [74] is more similar to ours in the sense that both questionnaire answers and measurements are acquired and alerts are sent when readings fell outside the previously established normal range. Also in [94] a monitoring system is reported that checks symptoms by means of a questionnaire and automatically checks for deviant answers. These studies all focus on general COPD care, whereas we focus on early detection of exacerbations. Also, by employing a Bayesian network we are able to perform more automatic data interpretation. Recently, an exacerbation risk alert-call system has been tested, based on weather forecasts, which was reported to be effective in predicting risk periods [44]. Another recent study on predicting exacerbations found that breathing frequency is a good predictor [111]. Currently their approach is limited because breathing frequency was measured via a home oxygen-therapy system. If a convenient way to monitor breathing frequency could be found, this would be a possible extension of our current system. For further references, see also the review of home care for COPD by Polisena et al. [79] and the telehealthcare review by McLean et al. [67].

3.7 DISCUSSION

We aimed to develop a clinically useful system to assist in patient management and in this paper have described the progress towards that goal. In this section we discuss limitations of the current implementation, lessons learned from the development, and we look at future work, extensions and possible generalisations.

The current version of our system has some limitations in the implementation that are a consequence of an iterative design process. We envision an autonomous system capable of disease management, but it is necessary to first construct and test the various components. Currently our probabilistic model functions as a classifier labelling the monitoring data as an exacerbation with some probability. We have shown that we can do this reliably, also on new data. To further assist in disease management we would like to infer the probability of health states ahead of time. In general chronic disease management can benefit from focussing on the analysis

of trends in time series. We are currently working on a temporal Bayesian network that extends the current model to take time into account.

Currently, we provide the patient with information about the likelihood of an exacerbation as a minimal kind of support. A logical next step is to analyse in more detail what kind of advice would be beneficial. In order to do so we will also take into account more background information about the patient, most notably current therapy plans. Patient feedback can then be extended to include adjusting therapy. So while the system we have now is mainly an alerting system – which is already an improvement compared to usual care – it will be extended to more specific advice. Clinical guidelines and input from clinicians is indispensable in formulating the right advice. Also more in general, for a decision support system to become successful it is important to have the cooperation of both patients and clinicians. As users of the system they will be able to comment on usability and suggest improvements.

These extensions are part of the route to a system that is capable of both autonomous functioning and more traditional alerting. Besides the improvements mentioned already, future work will include a larger evaluation on prospective data of model performance leading up to a randomised trial to evaluate the efficacy of the intervention. The current work however provided us with valuable insights about usability, by asking patients for feedback; allowed us to test the technical aspects of monitoring; and ensured that detecting exacerbations was possible. These findings form a good basis for the extended evaluation that is now being planned.

3.8 CONCLUSION

We have described the results of the development of a novel autonomous mobile assistance system for disease management, in particular focussed on COPD. The system described in this paper uses probabilistic reasoning to automatically interpret patient specific data to assist patients in managing COPD exacerbations. Initial testing shows that applying the system is technically feasible and patients are capable and willing to use the system. The model is well-founded on expert knowledge, literature and data, providing effective exacerbation detection. We have thus produced and evaluated on a pilot scale an advanced system architecture for home management of COPD exacerbations, with promising results. Future work will involve a more extensive test in a home-care setting, finally leading to a system capable of exacerbation detection in an early stage such that COPD exacerbation impact can be reduced and patients can self-manage their disease.

4

AUTOMATIC COPD-EXACERBATION PREDICTION: FINDINGS FROM A HOME-MONITORING PILOT STUDY

4.1 INTRODUCTION

Exacerbations of chronic obstructive pulmonary disease (COPD) cause considerable patient suffering [71] and regularly lead to hospital admission, resulting in significant health-care costs [108]. Prevention of exacerbations thus is a worthwhile goal, as has also been stated in global clinical COPD guidelines [40]. This is further supported by the finding that earlier treatment of exacerbations seems to lead to better outcomes [109]. It is therefore not surprising that mobile health (mHealth) is seen as a promising approach to improve COPD home-care and self-management. Our research fits within the general trend in health-care to develop eHealth solutions intended to improve care while reducing costs by means of computerised support systems. With our system we aim to reduce the impact of exacerbations on patient's well-being and the burden these acute events put on the health-care system. The main advantage of mobile monitoring is improved information provision, as the monitoring system can keep patients informed about their own health status and present data to health care professionals, which are not available in usual care, to keep them updated about their patients' health status. Moreover the system can assist in self-management by automatically giving the patient appropriate advice based on the monitoring data, which is usually not the case in existing telehealth solutions for COPD [67].

In the current article we report on a pilot study of a mobile COPD monitoring system for the detection of exacerbations at an early stage. A probabilistic model is used for automatic data interpretation. Data was gathered in the patient's home environment by means of a smartphone-based system that included a spirometer and pulse-oximeter. Predictions are based on the combined data from the sensors and the information on respiratory symptoms and signs acquired through an on-screen questionnaire on the smartphone display. The current system is a prototype and not yet optimal with respect to usability. A detailed description of the system can be found elsewhere [97]. An earlier version of the model has been developed in cooperation with pulmonologists from the Radboud University Medical Centre Nijmegen, using data from the TEXAS study [12] and validated using data from a subset of the London COPD cohort [51]. However, in the TEXAS study registrations took place biweekly, which is too infrequent for our current purposes. Also, not all variables that are part of our model were available in either data set. We now gathered new data with our home-monitoring system to assess the performance of the interpretation model for exacerbation detection and to assess whether the target population of elderly patients with COPD would be capable and willing to use such a system at home. We report on the monitoring data and model predictions of typical patients with COPD who either developed an exacerbation or recovered from an exacerbation.

Since the last registration, did you have dyspnea worse than usual?
Since the last registration, did you have more sputum than usual?
Since the last registration, did you have sputum of a different colour than usual?
Since the last registration, did you cough more than usual?
Since the last registration, did you have more wheezing breath than usual?
Since the last registration, were you limited in performing day-to-day activities due to COPD?
Since the last registration, did you have a feeling of malaise?
Since the last registration, did you have a fever?

Table 4.1: Questionnaire items (translated from Dutch).

4.2 METHODS

4.2.1 *Study design*

In this prospective observational pilot study patients with a confirmed diagnosis of COPD were instructed to use the monitoring system consisting of a smarthphone, a pneumotachograph micro-spirometer (Twente Medical Systems international, Oldenzaal, the Netherlands) and a pulse-oximeter (Nonin Medical, Amsterdam, the Netherlands) for a duration of four weeks. Measurements took place daily, at a prearranged time of the patient's preference, initiated by an automatic alert from the smartphone. The registration consisted of a questionnaire, answered via the smartphone's touchscreen, of 8 binary queries on whether patient's symptoms were worse than normal (Table 4.1). The questions are derived from questions used in the TEXAS study [12], which in turn were based on the Anthonisen criteria [5]. In addition to the questionnaire FEV_1 (Forced Expiratory Volume in 1 second) and SpO_2 (Saturation of hemoglobin with Oxygen as measured by pulse-oximetry) were measured. Patients were instructed by a nurse on how to use the system and how to perform the spirometry. The results of the sensor measurements were wirelessly transmitted by a Bluetooth interface to the smartphone. Data was automatically synchronised to a central server via a secured connection.

The system is capable of data interpretation by means of an advanced probabilistic model, which can compute the probability of an imminent or regressing exacerbation based on current data as provided by the patient. The model allows us to detect exacerbations automatically and based on the model prediction the system can provide direct feedback to the patient via the smartphone and to care providers via the secured web-interface. For this study we decided not to provide automatic advice to the participants based on the data, as the model is still in the process of being validated.

At the end of the four week period, patients were asked to fill in the System Usability Scale (SUS) [16], to assess usability of the home monitoring system. The SUS is a questionnaire consisting of 10 Likert scale items and is an often used reliable indicator of overall usability [9]. We used a version that was translated to Dutch. Three questions were added on the use of sensors and the smartphone touchscreen to obtain specific usability information for our system. The questionnaire with the additional questions will be referred to as the 'extended scale'.

Permission for the study was given by a medical ethics committee (CMO Arnhem-Nijmegen, the Netherlands; approval number 2011/242).

4.2.2 *Study population*

Ten patients with COPD were recruited from hospitals and general practices in the eastern part of the Netherlands. All participants gave written informed consent. Inclusion criteria were moderate to severe airflow obstruction (i.e. GOLD grades II or III) and sufficient cognitive capability to operate the monitoring system. We aimed to include two subgroups of patients: patients that were currently stable, but had experienced exacerbations in the past, and patients that consulted a physician because of a current exacerbation. Analysis of the first subgroup provided information on stable patients, who are 'at risk' for developing an exacerbation and might develop an exacerbation during the monitoring period; analysis of the second subgroup provided information on exacerbation recovery and possibly relapse.

4.2.3 *Data analysis*

The responses to the respiratory symptoms questionnaire were summarised into a single symptom score by counting the number of positive responses (i.e. indicating worse than normal). For FEV_1 the best result of three attempts was taken as the real value except where this resulted in implausible values ($FEV_1 > 6L$), which were considered measurement errors.

A dynamic Bayesian network model [69] was used for the model predictions. A dynamic Bayesian network models an uncertain process over time, which allows us to predict the value of a variable given observations of other variables. Here we use the network as a classifier that gives the probability of an exacerbation as output. Bayesian networks are robust against missing values, imputing them with the conditional distribution given the other observations or the prior distribution if there are no other observations. Predictor variables included in the model were the following: dyspnea, sputum volume, sputum purulence, cough, wheeze, activity, general malaise, fever, SpO_2 and FEV_1. The outcome variable of the model was exacerbation. However, in general, Bayesian networks are not limited to computing an outcome from predictors but also allow computing arbitrary probabilities of variables in the model given observations. For example, it is possible to compute the probability of observing certain signs given the observed symptoms. Details of the model will be published elsewhere [101].

Exacerbations were defined according to a slight adaptation of the symptom-based definition used by Bischoff et al. [12]: "A change for at least two consecutive days in two or more major symptoms (dyspnea, sputum purulence and sputum volume) or any one major symptom plus any one or more minor symptoms", where we defined minor symptoms as any symptom that was measured and was not a major symptom. The optimal parameters of the model were computed using this definition as 'ground truth', i.e. data records that satisfied the definition were considered an exacerbation for the purpose of estimating model parameters. The prediction probability threshold for exacerbations was computed via a receiver operating characteristic (ROC)-analysis [112]. Exacerbation events are predicted based on the observations from the last two days and are reported as the probability of the event for the next day. The situation is illustrated in Figure 4.1.

Time course plots were produced to compare the symptom scores to the model predictions and visually assess model performance over time. For each day during the monitoring period the symptom score is plotted together with the model prediction for the same day, which is based on the data from the previous day. The model prediction is the conditional probability of an exacerbation given the findings $P(E = \text{yes} \mid \text{findings})$.

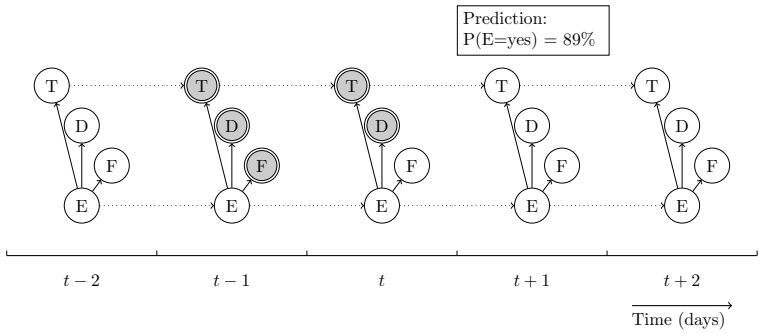


Figure 4.1: Part of the network model used for exacerbation predictions. Each time point represents a day. Variables shown are D = Dyspnea; T = Temperature; F = FEV_1 and E = Exacerbation. Shaded nodes indicate observations. The prediction shown is the probability when the presented observations are deviant from normal.

Patient's usability scores were computed following the standard procedure for the SUS, which results in a normalised score in the range 0-100 (higher is better). For the extended scale, scores were recomputed including the extra items and normalised to the same score range. Interpretation of the scores follows the analysis by Bangor et al. [9] to correct for score bias. The minimal score at which the system is considered acceptable is 70.

4.3 RESULTS

4.3.1 *Study population*

The ten patients with COPD that were enrolled had the following characteristics: 7 male and 3 female, between 53 and 76 years of age (mean (SD): 65.6 (6.8)) and FEV_1 % predicted $< 80\%$. Five patients were recruited while stable. One patient (from the stable group) withdrew participation due to perceived difficulty in using the system, caused mainly by a technical malfunction. The service interruption also caused missing values for other patients. In total 189 registrations were collected. Nine patients experienced an exacerbation according to the definition, with a total combined duration of 60 days.

4.3.2 *Model predictions*

The dynamic Bayesian network model predicts the onset of an exacerbation on the next day based on the current observations. The optimal discrimination threshold computed from the ROC-curve as the point closest to the perfect prediction point is $P(E = \text{yes} \mid \text{findings}) = 33.7\%$. Data points with prediction probability higher than this threshold are considered an exacerbation. With this threshold the classification accuracy of the model predictions over all data records is 81% (bootstrapped 95% CI: 0.72 – 0.87), with a false positive rate of 14% (bootstrapped 95% CI: 0.07 – 0.33).

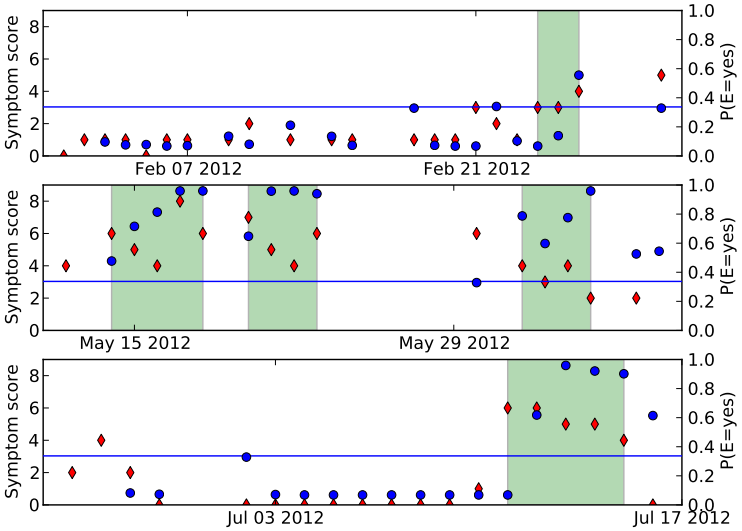


Figure 4.2: Time courses of monitoring data. Diamonds indicate symptom scores (left y-axis), circles prediction probabilities (right y-axis). The horizontal line delineates the classification probability threshold. Note that predictions are based on the observations from the two previous days.

4.3.3 *Time course analysis*

In Figure 4.2 time courses are shown that illustrate the performance of the model using symptom scores as a comparison measure. Three typical time courses are depicted: a stable patient that had a short increase in respiratory symptoms at the end of the time course; a patient with a prolonged increase in respiratory symptoms, indicating an exacerbation; and a patient that starts out with some symptoms, returns to a stable state and develops a sudden onset exacerbation. Predictions show a strong correlation with symptom scores, as expected. Sudden onset exacerbations are difficult to predict as can be seen in the first and third time course. A day-in-advance prediction will result in a false negative prediction and in these cases we should aim for early detection instead of prediction. In the first time course we observe a false negative prediction on the first day of the exacerbation, where the symptoms just reach the criteria for an exacerbation but FEV_1 is not affected. The second time course shows a more gradual onset, although due to the inclusion process of patients with current exacerbations the data is left censored. The model correctly classifies the start of the exacerbation and probabilities remain high during the course of the exacerbation, indicating correct predictions. Analysis of the third time course shows again that predicting sudden onset exacerbations is infeasible, however in this case the classification is correct on the first day of the exacerbation and the following days.

4.3.4 *System usability*

Eight patients completed the extended SUS questionnaire. The mean (SD) of the SUS scores was 67.2 (13.3), which according to Bangor et al. [9] is borderline acceptable. The extended score was 68.5 (12.9), which is not significantly different from the standard scale, suggesting that the smartphone touchscreen and the sensors did not adversely affect usability. It therefore appears that the system is reasonably usable for this group of relatively old patients, but improvements are possible.

4.4 DISCUSSION

Using an mHealth solution to assist in exacerbation management at home appears useful to limit the impact of exacerbations on patient health and potentially reduces health care costs by preventing use of acute care, out-of-hours medical services and hospital admissions. In this pilot study we tested the performance of an exacerbation prediction model to automatically interpret monitoring data with the aim to detect (relapse of) an exacerbation early. Model performance appears sufficient to assist in self-management, but more extensive research is required to establish this conclusively before the model can be applied to support patient care.

In a recent study a distinction was made between two exacerbation onset types [1]. Gradual onset and sudden onset exacerbations were identified, and in our data this divide is also visible. To successfully predict exacerbations it is important to take the type of exacerbation into account. Gradual onset exacerbation can be predicted more easily since there are measurable changes before exacerbation onset. For this type of exacerbation it is of interest to study in more detail the time course of the prodromal stage to see whether the dynamic Bayesian network can be optimised to predict this type of exacerbation. For sudden onset exacerbations, timely detection is more important but the sudden increase of multiple symptoms should be relatively straightforward to detect. The expert model may be useful to do so. The patient can also play an active role in this case by initiating a registration when their symptoms increase.

Our system captures relevant information to predict or detect exacerbations at an early stage which is a marked improvement over usual care where this information is usually not available, at least not from a mobile system and accompanied by automatic data interpretation [94]. Moreover, our initial usability test suggests that patients would be able and willing to use our home monitoring system. That the SUS score only reaches a borderline level of acceptability might be explained by the fact that the system used for this pilot study was a prototype lacking some polish and suffered from some technical malfunctions during the study period. These issues can be resolved and we expect that future versions of the system will have improved usability.

In our opinion, automatic interpretation of mobile monitoring data has great potential to detect at an early stage (relapse of) exacerbations in patients with COPD and improve self-management support, especially if combined with automatic feedback to the patient. The current model appears a useful step on the route towards an intervention that can be applied as a part of regular COPD patient care. This study was limited by the number of patients that was included. However, as proof of concept that automatic interpretation of monitoring data at home is possible, it provides sufficient support to warrant a follow-up study. In a validation study that is currently in preparation a larger group of patients from an in-patient pulmonary rehabilitation program will be monitored to obtain the required data to validate our dynamic Bayesian network.

ACKNOWLEDGEMENTS

We thank the patients that participated in the study; all health-care professionals that cooperated in carrying out the study; Riet Cretier for assistance with the logistics and Bas Lijnse for technical support. Data to support the development of the model were provided by John Hurst, Gavin Donaldson and Wisia Wedzicha from the London COPD cohort. Part of this work was supported by a grant from the Netherlands Organisation for Health Research and Development, ZonMW Project 300050003 and by a research grant from Boehringer Ingelheim.

5

LEARNING BAYESIAN NETWORKS FOR TIME SERIES ANALYSIS

5.1 INTRODUCTION

Clinical data often takes the form of time series, which, when interpreting all the variables concerned in their mutual context, offer a description of the progression of a disease over time. While insight into the evolution of a disease is an important aspect of the management of any disease, whether acute or chronic, for patients with a chronic disease the evolution is of even more importance, as often the disease will not disappear. In this context it is in particular important to be able to detect when the disease becomes worse, i.e. to detect and possibly prevent *exacerbations*. For any chronic disease it is therefore of clinical interest to study the interaction between different variables, ranging from signs and symptoms to environmental factors, in terms of both static and temporal relationships. If we can capture this knowledge in a model, predictions regarding health status made by means of such models can be used to assist in chronic disease management, for example by advising on therapy adjustment. Furthermore, disease models are of paramount importance for epidemiological purposes, for example for survival analysis, as well as for cost-effectiveness analysis and policy planning.

A complication that arises when analysing clinical time series is that it is often hard to obtain sufficient data, for example because the event of interest is relatively low frequency or because taking measurements is costly, time consuming or inconvenient for the patient. In addition, the reality of gathering clinical data is that observations are made at irregular time intervals and the data will contain missing values. Patients will sometimes forget to provide data, or omit some evidence for unknown reasons. Also measuring devices may sometimes fail or readings may not be recorded. These observations pose a challenging research question, which we seek to answer in this paper, namely, whether we can learn useful predictive models from clinical data with the combined characteristics of missing values and limited availability.

The research methods we propose draw their inspiration from various existing methods, which have proven to be successful in general machine learning applications. However the combination of methods has not been applied to clinical time series analysis. Temporal Bayesian networks are our main tools to reason about causal and temporal processes in a probabilistic manner. They provide interpretable and versatile models to describe time series data and can be used to classify states and make predictions about future states. To learn network structures in the presence of missing observations we make use of the structural expectation-maximisation (EM) algorithm [35], which iteratively completes the data and performs a search for the best network structure to explain the data. Here we employ a variant of structural EM, with a different approach to filling in values for the missing data. Finally, to tackle the problem of data sparsity, we consider bootstrap methods originally developed in statistics [32]. In the context of Bayesian networks these methods have

been applied to the analysis of micro-array data and we extend them here to the learning of temporal models from clinical time series.

To provide a concrete clinical context for this research, we focus on chronic obstructive pulmonary disease (COPD) as an application area. This disease has many characteristics that are typical for any chronic disease, although symptoms and signs will be mostly different from those other diseases. COPD is a major chronic disease in terms of morbidity and mortality; it affects the respiratory system, decreasing lung capacity and obstructing airways, thus interfering with normal breathing. An important aspect of COPD which is particularly relevant in the present context is the progressive nature of the disease. Specifically episodes of acute deterioration have a profound impact on patient well-being and on health-care costs [104]. These exacerbations are mainly caused by airway infections resulting in symptom worsening [108]. Important to note is also that patients with frequent exacerbations usually have faster disease progression, which makes exacerbation prevention a particularly relevant goal. Additionally, a faster treatment response to exacerbations leads to better recovery [109].

The main contributions of the paper are as follows:

- We formulate an algorithm to learn temporal probabilistic models from limited clinical time series with missing values. The main novelty of the algorithm lies in combining learning of dynamic Bayesian networks from clinical data using structural EM with block bootstrapping for small data samples.

- We propose a variant of our learning algorithm based on naive Bayes networks, which has the attractive properties of reduced computational complexity, thus easy construction, while offering good prediction performance.

- The proposed learning algorithms are used to build predictive models of COPD patient's health status, focussing on day to day progress of signs and symptoms that can rapidly change during exacerbation events. These predictive models are novel in the context of COPD as they can handle both the dynamic nature and uncertainty inherent in the disease progression. As such, these models can be embedded within clinical or home-based applications for chronic disease management.

- We evaluate the learning procedure on COPD synthetic and patient data and show that it is effective in terms of structure discovery of interesting variable relationships, interpretability and prediction performance of the models learned.

- The results from this research demonstrate important clinical implications not only for the prediction of COPD exacerbations but also for the clinical relevance of the methods proposed for chronic disease management applications in general.

5.2 RELATED RESEARCH

5.2.1 *Clinical time series analysis*

Data analysis is a typical statistics topic that plays a crucial role in medicine. Survival analysis [85] is a good example of time series analysis, but the standard technique of Cox-regression is however less flexible than Bayesian network techniques. Our work is more closely related to the research in [102], which argues for using dynamic Bayesian networks for prognostic models in medicine. Variants of dynamic Bayesian networks have also been used to model temporal dynamics of organ failure in patients in intensive care units (ICU) [77], although there no structure learning

was used. Further, a Bayesian network has been developed on the basis of electronic health record data to predict the onset of COPD in asthma patients, however, temporal information was not explicitly taken into account [49]. In the specific context of long term disease management for COPD, related research has focussed on facilitating remote communication [67, 79] and automatic data interpretation is still uncommon. In [10] a telehealth system is described that has been applied to COPD and contains a decision support component. However, currently the decision support is limited to rule based detection of abnormal values and trend detection. Automatic interpretation of monitoring data using machine learning while taking time and uncertainty into account therefore appears to be a useful contribution to the area of chronic disease management.

5.2.2 *Modelling techniques*

Early work on using Bayesian networks for prediction includes dynamic network models [26], with e.g. a clinical application to predicting sleep apnea [25]. We use and extend techniques for learning Bayesian networks from data with missing values [35] and learning from small samples using bootstrapping [37]. These methods are used extensively in bioinformatics [56], but application to the domain of clinical time series analysis constitutes a new and interesting challenge. The bootstrap methods used for small data samples are related to what is known as *bagging* in the machine learning literature [14], but are usually applied to learning decision trees instead of Bayesian networks. Our augmented temporal naive Bayes model is an extension of the TAN classifier from [36] to a prediction model that takes time dependencies into account. Other approaches to learning dynamic Bayesian networks include using steady state information [59] and introducing hidden variables to model process changes [95].

5.3 THE AERIAL PROJECT: MOBILE COPD MANAGEMENT

The methods we developed and use in this paper were needed as part of a research project, called Aerial, aimed at the detection of worsening in patients with COPD, i.e. exacerbations. Here we briefly describe the system and the design of the study to sketch the practical clinical context of the work.

5.3.1 *A system that supports self-management*

In order to facilitate self-management of COPD by patients we developed a system with the capability to gather patient-specific information, to interpret the gathered data automatically and to offer feedback and advice to the patient. The general architecture of the Aerial system is shown in Figure 5.1. The system consists of a smartphone as the main component taking care of communication and computation. Questionnaire data is collected from the patient on the smartphone, which also communicates wirelessly with the sensors used to obtain objective information on the patient's health status. A web-based system allows scheduling tasks and collecting patient data centrally. The web-centre receives the data from the smartphone and provides data access for health-care workers. Patient data are interpreted in the smartphone by means of a disease-specific probabilistic model, different variants of which are studied in this paper. A more comprehensive description of the system can be found elsewhere [96, 97].

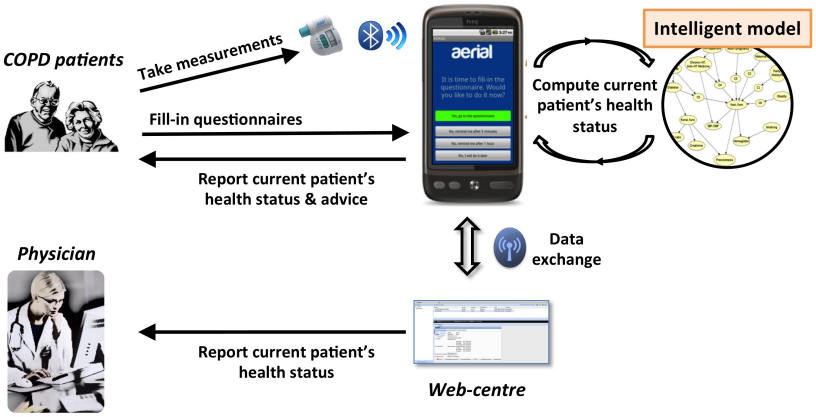


Figure 5.1: Schematic of the system setup.

5.3.2 *Patient population and data gathering*

The aim of the study was to investigate the prerequisites to facilitate self-management of COPD-patients in a home-care setting. We tested the Aerial system at home to see whether this would result in usable data with the goal of gathering a data set from which to learn models for COPD-exacerbation prediction. Ten participants were recruited from hospitals and general practices in the Netherlands, 7 male and 3 female, between 53 and 76 years of age (mean (sd): 65.6 (6.8)). All participants gave written informed consent. Inclusion criteria were GOLD II or III (disease severity classification on a scale from I to IV) [40] and sufficient cognitive capability to operate the system. There were two possible inclusion paths: stable patients, although having had exacerbations in the past, and patients that reported to a physician because of an exacerbation. The first path provides information on stable patients, and possibly on exacerbation onset; the second on exacerbation recovery and possibly relapse.

Patients were monitored daily for a duration of approximately 4 weeks. Each day they answered a set of binary questions about their symptoms and performed spirometry and pulse-oximetry measurements. All questions were formulated such that a 'yes' answer indicates a symptom worse than the baseline condition for that patient. The set of recorded variables consists of the symptoms dyspnea (D), sputum volume (SV), sputum purulence (SP), cough (C), wheeze (W), temperature (T), malaise (M) and activity (A); plus the measurements SpO_2 (SO) (blood oxygen saturation) and FEV_1 (F) (forced expiratory volume in 1 second). Exacerbations (E) were defined using a clinical symptom-based definition (see e.g. [12]). A total of 189 data records were collected, which when regularised by adding missing values for days that were not recorded resulted in a data set of 250 records. Of these records 60 were during an exacerbation. Missing data was partly a consequence of technical issues and partly due to patients omitting data for unknown reasons.

5.4 PROBABILISTIC MODELS FOR TEMPORAL DATA ANALYSIS

As the research aim was to develop models that were able to interpret temporal clinical data, in particular revealing the progression of COPD including its stochastic variation, especially *probabilistic models* were considered to be attractive for that

purpose. Probabilistic models can express temporal trends and handle missing values when data of a specific patients must be interpreted. The successful use of dynamic Bayesian networks in the context of micro-array analysis [56], where sparsity of data is also a problem, made us wonder whether similar methods might also be successfully applied to clinical problems.

We start with a brief summary of basic methods used in the research, leading up to a method to learn models from sparse clinical time series data.

5.4.1 *Dynamic Bayesian networks*

A *Bayesian network* [23, 75], is a probabilistic graphical model represented as a pair $\mathrm{BN} = (G, P)$. Here, $G = (V, A)$ is a directed acyclic graph consisting of *vertices* V, corresponding one-to-one to random variables of interest, and $A \subseteq V \times V$ are *arcs*, representing dependencies between variables. Furthermore, P is a joint probability distribution defined by a family of conditional probability distributions of the form $P(V \mid \mathrm{pa}(V))$, that is, the probability that V takes on a specific value given the values of its parent variables, $\mathrm{pa}(V)$. The network represents the joint distribution over the random variables, which can be factored according to the dependences represented in the graph, resulting in:

$$P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P(V_i \mid \mathrm{pa}(V_i)),$$

where $V_i \in V$ is the representation of a random variable in the graph G. Any probability of interest can be computed from this joint probability distribution.

A *dynamic Bayesian network* [28, 69], DBN for short, is an extension of a Bayesian network to a distribution over a sequence of random variables. It is particularly well suited to represent a Markov process

$$X_1 \to X_2 \to \cdots \to X_t \to \cdots$$

where X_t represents a random variable at a particular moment in time t. The joint distribution can be decomposed using the chain rule, writing $X_{1:T}$ for $X_1, \ldots, X_t$, $\ldots, X_T$:

$$P(X_{1:T}) = P(X_T \mid X_{1:T-1}) P(X_{1:T-1}).$$

In general a DBN is a factorisation of a probability distribution, like its atemporal counterpart. In addition, each X_t can be represented as a BN. This BN is called a *time slice* and relations between time slices can be modelled by introducing arcs in the graph between random variables in different time slices. A DBN factorisation can then be written as:

$$P(X_{1:T}) = \prod_t \prod_i P(X_{t,i} \mid \mathrm{pa}(X_{t,i}))$$

where i indexes variables within a time slice and $\mathrm{pa}(X)$ denotes the parents of X in the graph. A hidden Markov model, a popular stochastic model used for pattern recognition, is a special case of a DBN.

A common assumption is that there is only a limited time frame that influences the current state of the process, as opposed to the complete history, which simplifies model learning. When assuming an nth-order Markov process we obtain:

$$P(X_{1:T}) = P(X_T \mid X_{T-n:T-1}) P(X_{T-n:T-1}),$$

recursively. In the context of clinical data analysis, this assumption makes sense for two reasons: first, as time passes physiological and disease processes will change and older information will be less informative about the current state of the patient; second, it will often not be possible to obtain reliable information about the past. For smaller n more temporal independence is introduced and when only sparse data is available, it is common to make the most restrictive version of the Markov assumption, first-order, such that the future state of the process only depends on the present:

$$P(X_{t+1} \mid X_{1:t}) = P(X_{t+1} \mid X_t).$$

Hence, all parents of a variable X will be in the same time slice or in the previous time slice. From a medical point of view this means that the current health status provides the most information about the future. Given that clinical data is often sparse, an important practical consequence of this assumption is that it simplifies the model, and hence reduces the amount of data we need. When we now also assume that the process is *stationary*, that is $P(X_{t,i} \mid \mathrm{pa}(X_{t,i})) = P(X_{t',i} \mid \mathrm{pa}(X_{t',i}))$ for all t, t', we obtain a *two-slice DBN* consisting of an initial network BN_0 and a transition network $\mathrm{BN}_{\rightarrow}$. A process through time can now be modelled by a sequence of repetitions of transition networks. When modelling a chronic disease over a long period, it may be that stationarity is not a reasonable assumption. For COPD one might argue that it is also useful to consider separate models for the disease stages (GOLD I-IV [40]). The COPD data we study here is, however, limited to GOLD II and III by the inclusion criteria, and the amount of available data is too limited to make a distinction. In general however, it appears useful to consider recent techniques to learn non-stationary DBNs [81].

5.4.2 *Model learning*

Given a data set we can use machine learning techniques to learn a model from the data. Two main tasks are usually distinguished: (*i*) finding a network structure, and (*ii*) finding the parameters that best describe the data given a network structure. See [57] and [68] for fairly comprehensive overviews of probabilistic learning.

Parameter learning

Parameter learning entails finding the optimal parameters $\hat{\theta}$ for a given network structure G, that explain the data D:

$$\hat{\theta} = \operatorname*{argmax}_{\theta} P_G(D; \theta).$$

This is the maximum likelihood estimate, parametrised by θ, which is straightforward to compute. It may suffer from overfitting, especially for sparse data, which is the case that we are considering here. A Bayesian approach makes explicit the uncertainty in the parameters by taking θ to be a random variable, leading to:

$$P_G(\theta \mid D) = P_G(D \mid \theta) P_G(\theta) / P_G(D),$$

from Bayes' theorem, where we compute a distribution over parameters, with some prior distribution $P_G(\theta)$. Assuming independent and identically distributed data, global and local parameter independence, the likelihood can be decomposed in local likelihood terms according to the graph:

$$P_G(D \mid \theta) = \prod_D \prod_{i=1}^{n} P_G(X_i = k \mid \mathrm{pa}(X_i) = j, \theta).$$

Since we are using discrete variables in our COPD model, we use multinomial distributions for the local likelihood terms $P_G(X_i = k \mid \mathrm{pa}(X_i) = j, \theta)$. The conjugate prior distribution of the multinomial is a Dirichlet distribution, where conjugacy implies that the posterior is also a Dirichlet distribution. We can interpret the hyperparameters α of the Dirichlet distribution as pseudo-counts which leads to a parameter estimate

$$\theta_{ijk} = P_G(X_i = k \mid \mathrm{pa}(X_i) = j) = \frac{|D[k,j]| + \alpha[k,j]}{|D[j]| + \alpha[j]},$$

where $D[k,j]$ (and analogously for α) is an index expression selecting the data cases where $X_i = k$ and $\mathrm{pa}(X_i) = j$ and $D[j] = \sum_k D[k,j]$.

Parameter estimation using likelihood decomposition properties works well for complete data, which is unrealistic for clinical models. The expectation-maximisation (EM) algorithm can be used to learn parameters when we have missing values [31].

Let $D = \langle O, H \rangle$ be a data set consisting of observed values O and hidden values H; O_i, H_i indicate a single data point and $\tau(D) = \sum_i \tau(O_i, H_i)$ the sufficient statistics. The EM algorithm iteratively adjusts the parameters in two steps:

$$\text{E-step: } \mathbb{E}[\tau(D)] = \sum_i \mathbb{E}_{P(H_i \mid O_i, \theta^s)}[\tau(O_i, H_i)]$$

$$\text{M-step: } \theta^{s+1} = \operatorname*{argmax}_\theta P(D \mid \theta)$$

where s is the step counter. The E-step completes the data based on the expected values for the hidden variables given the current parameters and the M-step computes the maximum likelihood parameters based on the completed data. This algorithm provably improves the likelihood in each step until some maximum is reached, although this could very well be a local maximum [31]. EM gives us no guarantee of finding a global maximum; however, its results are normally sufficiently good.

Structure learning

Hand-crafting a network structure is difficult and time consuming and hence it is useful to also try to learn the network structure from data. Even for a limited number of variables as we have in our case, the search space is of such a size that exhaustive search is infeasible. Alternative methods include constraint based and score based techniques. See [2] for a comparison on a large set of hospital management data, where constraint based methods appear to perform somewhat better. However, an important disadvantage of constraint based methods, that rely on independence tests, is that early errors propagate and can have large effects. Additionally, independence tests tend to be less reliable especially on small samples, for example because distribution assumptions are not met. In [27] arguments for and against constraint based methods are examined and a hybrid approach is proposed, which is however not directly applicable to time series. We opt here for a standard score-based search procedure that tests local changes to the graph, greedily maximising the score $\log P(G \mid D)$, which by Bayes' rule can be written as

$$\log P(G \mid D) \propto \log P(D \mid G) + \log P(G).$$

Taking a Bayesian approach the probability $P(D \mid G)$ can be written as the integral

$$P(D \mid G) = \int_\theta P(D \mid \theta, G) f(\theta \mid G)\, \mathrm{d}\theta,$$

with $f(\theta \mid G)$ a probability density function, marginalising out the parameters. For complete data the likelihood again decomposes according to the graph and assuming

conjugate priors there is a closed form solution for the marginal likelihood $P(D \mid G)$ (see for example [57]). A common choice for the structure prior $P(G)$ results in the BDe-score [47], which uses a prior such that we have score equivalence. That is, networks that encode the same independence information will have the same score.

Unfortunately, the task becomes increasingly difficult with missing data. Finding optimal parameters when we have missing values depends on the structure, while structure search clearly depends on the data, including the missing values. There is no particularly efficient solution to this problem, but we can use the same general principle as we did to learn parameters from partially observed data. This idea is called structural expectation maximisation [35]. In essence the structural EM algorithm alternates between completing the data based on the current structure and parameters and finding a model that has a better score. The same caveat as mentioned for the EM algorithm applies here as well, structural EM does not guarantee finding the global maximum likelihood solution. When the number of missing values increases the search space becomes larger and we are less likely to find a globally optimal model.

5.4.3 *Small data samples*

With the model learning techniques described above we can try to find models that reasonably fit our data, even when some of the data is missing. However, as we are learning statistical models there is a strong dependence on the amount of available data. The more data we have the better we can learn relations. Unfortunately, gathering data is often a difficult, time consuming process. For the COPD self-management application the demanding logistics of studying a new system in a home-care setting, made data collection hard. Furthermore, sometimes the model space is so large that obtaining sufficient data to learn models directly may be infeasible. Previous research in determining gene interaction patterns from micro-array data [37, 56], is another example of research that stumbled across the mentioned problems. The clinical data we study here is, however, of a different nature than micro-array data, as signs and symptoms will clearly produce other temporal dynamics. Although the number of variables was fairly limited in our case, we had a combination of missing values within records and a limited number of records in total. Hence we needed to combine the EM procedure with some technique to deal with the small data sample.

BOOTSTRAPPING The data sparsity in modelling gene data led to using bootstrapping [32] as a way to estimate the uncertainty of relations in Bayesian networks [37]. The idea is as follows: bootstrapping can be used as a nonparametric estimate of a statistic of interest, for which we can take the presence of arcs in the Bayesian network graph. The intuition behind bootstrap replications of a data set is that if we assume a generative model, the actual data we see are just a particular realisation which may be unrepresentative of the underlying process. Bootstrapping a number, say m, of new realisations by resampling (with replacement) $|D|$ samples from the original data and learning models from those data sets, we can estimate whether an arc a should be present in our model:

$$P(a) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}(a, G_i),$$

where $\mathbb{I}(a, G_i)$ is an indicator function that is 1 when a is present in the graph learned from bootstrap realisation i.

Since we are analysing temporal data, the situation is somewhat more difficult. Simply resampling from the original data will discard all the ordering information,

which is clearly undesirable. Instead we have to apply a kind of bootstrapping that preserves the correlation between different time points. In the statistics literature a number of methods have been developed to do so, and we opted to use fixed-length block bootstrapping. As the name suggests this entails resampling blocks of data points, which preserves the time relations within the block (see e.g. [60] for a theoretical comparison of block bootstrap methods). We used the implementation from the R-package *boot* [17], with a fixed block-length of 5.

From the bootstrap results we can construct a model that includes the features (arcs) that have a high probability. Another possibility is to use model averaging, where we compute the prediction probability on validation data D_{val} as the average over the predictions of the bootstrap models

$$P(D_{\text{val}}) = \frac{1}{m} \sum_i P(D_{\text{val}} \mid G_i). \tag{5.1}$$

However, each of the networks G_i has some score attached to it, so a better approximation to the real probability can be obtained from the weighted average

$$P(D_{\text{val}}) = \frac{1}{m} \sum_i P(D_{\text{val}} \mid G_i) P(G_i \mid D_i). \tag{5.2}$$

5.4.4 *A method to learn temporal models from small data samples with missing values*

With all the machinery described above, we can now start putting together the method to learn models from time series. The main idea is a variant of structural EM applied to bootstrap resampled data. The new features of this procedure are that it combines sparse data techniques with structure learning on small samples of time series data, which is typical for clinical data.

Algorithm 1:

```
1 Construct m data sets D_{1:m} from the original data D using block
  bootstrapping
2 foreach D_i  i ∈ {1, ..., m} do
3 |   for j = 0, 1, 2, ... until convergence or reaching a time limit do
4 |   |   Learn parameters θ_j for the DBN structure G_j with EM
5 |   |   Complete the data D_i by sampling from the posterior P(H | O, θ_j)
6 |   |   Search for a DBN structure G_{j+1} on the completed data that improves
  |   |   P(G_j | D_i)
7 |   end
8 end
9 Compute the model average predictions
```

This algorithm generalises the bootstrap method for sparse data [37], to temporal models. In order to learn the structure of a temporal model, we need a generalisation of the structural EM algorithm [38]. In each iteration we learn the structure of a two-slice network consisting of BN_0 and $\text{BN}_{\rightarrow}$. There are however some differences with the algorithm in [38]. In particular, we take a different approach to completing the data. On line 5 of our algorithm, we sample values from the posterior distribution $P(H \mid O, \theta_j)$ instead of computing fractional sufficient statistics. That is we sample values for the missing data H from the posterior given the observed values

O and the current parameters θ_j. This allows us to decouple the parameter learning and the structure search. The result is a complete data set without (fractional) expected values filled in for the missing values, which lets us use structure search methods developed for complete data. Additionally, sampling from the posterior $P(H \mid O, \theta_j)$ has the interesting property that it results in automatic small data perturbations. Because structural EM is not guaranteed to find a global maximum, the data perturbations help in starting the search from slightly different departure points. The posterior sampling provides a way to let the search reach different parts of the search space. In [54] some of the differences between hard-assignment, EM (soft-assignment) and posterior assignment are studied in the context of clustering, which is closely related to filling-in missing values.

Some remarks on implementation. The EM parameter learning is implemented using Smile [30], and the DBN structure search from the completed data is implemented in Banjo [45]. In practice, convergence of the complete procedure can be very slow because the search space is large. Within each iteration we use the generalised EM result that it is sufficient to update the model when it has a higher score, without having to maximise the score. Per iteration we set a time limit on the structure search only accepting the new structure when a higher score was found.

AUGMENTED TEMPORAL NAIVE BAYES Although a full model is useful to gain insight into the domain, for our application it is important that the performance on predicting exacerbations is good. It therefore appears useful to emphasise this goal during model construction. To do so we start from the concept of a naive Bayes classifier, with *exacerbation*, E, as class variable. The rationale is that naive Bayes is a good baseline classifier that can be extended with the information obtained from structure learning. The idea of modelling dependencies between naive Bayes feature variables has been used to construct tree augmented naive Bayes (TAN) classifiers [36]. We here propose an augmented temporal naive Bayes classifier. The presence of arcs from our class variable *exacerbation* to each of the other variables in the same time slice is enforced, ensuring the naive Bayes structure. Structure search can then identify dependencies between the other variables, also through time. We retain the inner loop of Algorithm 1 to find the dependencies that best explain the data, but instead of bootstrapping we restrict the model search space by predefining part of the network structure. The acyclicity of the graph implies that $X_t \rightarrow E_t$ is not present, whereas we need to explicitly blacklist arcs of the type $X_t \rightarrow E_{t+1}$ for all X. The structure search will identify temporal dependencies but one could argue that since we are interested in the temporal behaviour of the class variable that we should also enforce the presence of the arc $E_t \rightarrow E_{t+1}$. It turns out that in this specific case of a COPD model, when the arc is not enforced, it is found by the search procedure.

5.5 EXPERIMENTAL SETUP

In the previous section we described a general method to learn dynamic Bayesian networks from small sample time series data. Here we describe the experimental evaluation of the methods for the clinical problem of predicting exacerbations of COPD. The experiments serve to evaluate, first, whether the models that result from the learning procedure are accurate; second, to explore what models can be learned from the COPD data. For the first goal we use synthetic data from a synthetic model, described below. The second goal can be further refined, as we are interested in how the learned models compare to the model constructed in cooperation with a pulmonologist; and we want to find out what the performance of these models is on predicting exacerbation events.

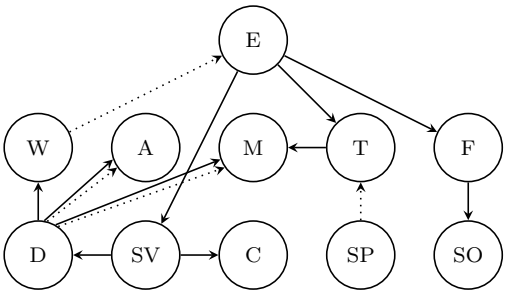


Figure 5.2: Model used to generate the synthetic data. The variables are A = activity; C = cough; D = dyspnea; E = exacerbation; F = FEV_1; M = malaise; SO = SpO_2; SP = sputum purulence; SV = sputum volume; T = temperature; W = wheeze.

5.5.1 *Data*

Synthetic data

To test our model learning procedure we first study the results with synthetic data from a known model. The model contains the same variables as the Aerial data (Section 5.3) to stay as close as possible to the real data context. The arcs in the model were chosen such that the relations are intuitive, but were not meant to be clinically correct per se. Analogous to the other models both atemporal and temporal arcs are in the model. The structure of the model is shown in Figure 5.2. The parameters were assigned by hand, without attempting to capture the real relations. From this model we generated four data sets of the same length as the Aerial data, one without missing values and the others with 10%, 20% and 30% missing values respectively. The latter is similar to the percentage of missing values in the Aerial data. We then applied our model learning procedure to each of these data sets in order to be able to compare the learned models to the known source model.

Aerial data

The Aerial data consists of time series from 10 patients, gathered during the pilot study of our disease management system. The characteristics of this data set have been described in Section 5.3 as part of sketching the background for this work. The data set is used as the input for the learning procedure, to learn models that explain the COPD-monitoring time series.

Validation data

An independent data set [51] was used for validation. It provides time series of COPD-exacerbation related variables, used solely for external validation, so no model parameters were learned from this set. As the data is used for testing our model retrospectively, the set of variables is incomplete. The data consists of time series from 13 patients of the London COPD cohort who had an exacerbation, with a total of 2849 data entries, of which 406 were during an exacerbation. The data contains values for the variables dyspnea, sputum volume and purulence, wheeze, cough, temperature, and oxygen saturation (SpO_2). We consider two variants of this data set, the complete validation set, denoted by D_{val}, and a deduplicated version D_{dedup}. The latter consists of the same data, but with consecutive identical entries removed. The idea behind removing duplicates is that we are interested in

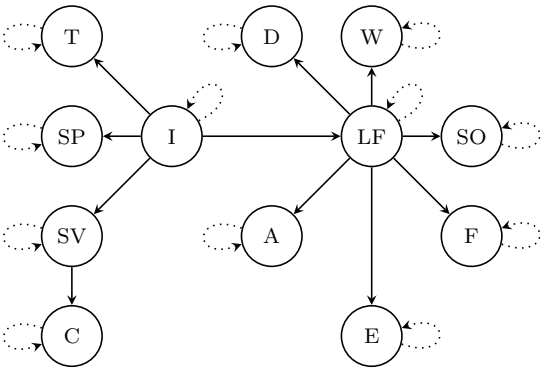


Figure 5.3: Expert model, with added temporal identity arcs. The variables are A = activity; C = cough; D = dyspnea; E = exacerbation; F = FEV_1; I = infection; LF = lung function; SO = SpO_2; SP = sputum purulence; SV = sputum volume; T = temperature; W = wheeze.

predicting relevant state changes instead of finding models that are often correct by predicting that nothing changes. By removing repeating sequences we ensure that we make predictions with data that frequently changes state. Note that we only remove exact duplicates, so a change in the observations without a change in exacerbation label still constitutes a change.

5.5.2 *Expert model*

The model constructed in cooperation with pulmonologists from the Radboud University Medical Centre, described in [96, 97], is used as a baseline model. This model is static, in the sense that it does not explicitly model time effects. To make temporal predictions we simply interpret the exacerbation probability as the prediction probability at a future time point. Additionally we constructed a dynamic version of the expert model by adding identity arcs to each variable. The graph is shown in Figure 5.3. For the additional parameters we choose values heuristically, assuming that remaining in the same state is more likely than switching state. It should be noted that although we refer to this model as dynamic expert model, the pulmonologist was only involved in the construction of the static model and the dynamic version should therefore be seen only as a naive extension for comparison purposes.

5.5.3 *Evaluation metrics*

To evaluate the models and their performance we will use standard metrics from classification. Classification evaluation is often based on measuring true positives (TP), cases that are correctly classified as positive; false positives, incorrectly classified as positive; and analogously for true/false negatives (TN; FN). The true positive rate (TPR) is then defined as $TP/(TP+FN)$ and the false positive rate (FPR) as $FP/(FP+TN)$. Plotting a curve of TPR-FPR for different cut-off points results in an ROC-curve, which is often summarised in a single number by computing the area under the curve or AUC.

We distinguish two situations: (*i*) the performance of network structure discovery and (*ii*) the prediction performance of different models. For structure discovery, a true positive is an arc present in both the original and learned network graph and a

Data set	TP	FP	FN	TPR
Complete	11	4	3	$11/14 \approx 0.79$
Missing 10%	9	4	5	$9/14 \approx 0.64$
Missing 20%	8	3	6	$8/14 \approx 0.57$
Missing 30%	4	5	10	$4/14 \approx 0.29$

Table 5.1: Structural equivalence scores for models from synthetic data.

false positive is an arc not present in the original network. True and false negatives are defined analogously. For prediction, the usual interpretations in terms of data records is used.

To evaluate the prediction results we also compute the Brier score [15], which is defined as $\frac{1}{|D|}\sum_{i=1}^{|D|}(p_i - l_i)^2$, where p is the predicted probability and l the correct label. A Brier score of zero indicates perfect prediction.

5.6 RESULTS

5.6.1 *Synthetic data*

To get a feel for the performance of the methods we used, let us first look at the results of synthetic data from a known model (Figure 5.2). The data are similar to the real data, so the performance of our learning methods should tell us something about the performance on the real data. Table 5.1 shows the true positives and false positives/negatives of the learned model arcs. The true negatives are less interesting because the learning procedure is forced to only consider graphs with bounded indegree. For the complete data we do not actually need the structural EM method, but for completeness we still report the results.

Some interesting observations can be made. Looking at the true positive rate (TPR), we see that as the percentage of missing values increases the TPR decreases, as expected. True and false positives do not tell the whole story however. It is not in general possible to establish the direction of arcs based only on observational data, hence we may learn networks with arcs that are reversed. With simple scoring this results in both a false positive and a false negative for a single reversed arc. This occurs in the model with 10% missing values where the directions of the arcs on the path $E \to SV \to C$ are reversed and in the model with 30% missing values where the arc $E \to SV$ is reversed. Also, failing to detect that there is both a direct and a temporal arc $D \to A$ and $D \to M$ is not a very serious issue, because whether these should be distinguished as separate arcs is quite sensitive to the parameters. Furthermore, all models missed the arc $F \to SO$, but looking at the parameters we see that the distribution is close to uniform which makes it almost impossible to detect the dependence from limited data.

We also computed bootstrap models for the data with 30% missing values. In Figure 5.4 the resulting aggregated model is shown. For each bootstrap replication we computed the best network using the procedure described above and then we averaged over the models to compute the probability of finding the arc. Due to the computational cost of the search procedure, bootstrapping was restricted to 9 resampled data sets. The arcs shown in the graph were found in a majority of the bootstrap models, i.e. the probability of the arc is $P(a) > 0.5$. Because the scores of the best networks are fairly close together, it makes little difference whether we use Equation (5.1) or Equation (5.2) to compute the arc scores. Observe that there are only a limited number of arcs with a high score, which is to be expected if there are multiple models that explain the data reasonably well. The arcs that are found are

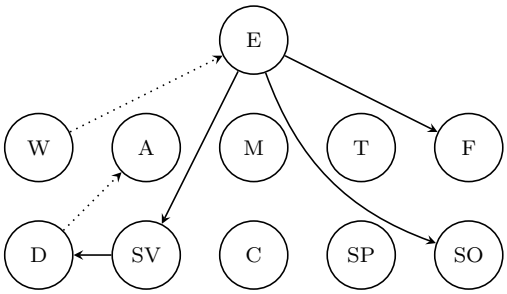


Figure 5.4: Majority vote of arcs according to the bootstrap models.

correct, except for $E \to SO$, which means 5 arcs are true positives and 1 is a false positive. The TPR is $5/14 \approx 0.36$ which is somewhat better than the result without bootstrapping for data with 30% missing values.

As usual, structure learning has the twofold goal of elucidating the structure of the domain and performing density estimation. It turns out that the bootstrap method allows us to identify important edges in the graph, which serves to fulfil the first goal. We now turn to the goal of density estimation, for which prediction performance can be used as a measure. In the context of COPD management, we are primarily interested in the probability of exacerbation. We compute the probability $P(E_{t+1} \mid O_t)$, which is the prediction probability of an exacerbation at time $t+1$ given the observations at time t (recall that the observations were made daily, so a single step is a day). A new data set with 30% missing values was generated from the original synthetic model as test data. An ROC-analysis shows that all models perform almost identically, with an area under the curve AUC = 0.63. This performance may seem low, but an ROC-analysis with the true model that generated the data also results in an AUC of 0.63. This is a consequence of the fact that the generating model contains limited temporal dependencies. Computing the autocorrelations for the generated time series also clearly shows that correlations are small. Therefore we cannot expect to make very good predictions on these data. Inspection of autocorrelations in the real data set shows the presence of more temporal information, so prediction performance analysis will be of more interest there.

The analysis of synthetic data indicates that even from a small sample of time series data we can reasonably reconstruct the network structure. As expected the performance is sensitive to the presence of missing values, but it appears that some amount of missing values is tolerable. These results provide sufficient confidence to analyse the COPD-monitoring data.

5.6.2 *Aerial data models*

The Aerial project aims to provide disease management for COPD patients. In order to achieve this goal we have performed a pilot study as described in Section 5.3. We now turn to the analysis of the gathered data to see whether our learning procedure is capable of finding relevant patterns.

Model learning and bootstrapping result

The result after 100 iterations of structural EM is shown on the left in Figure 5.5. We will refer to the model learned from the Aerial data as *Model A*. It is not as easy

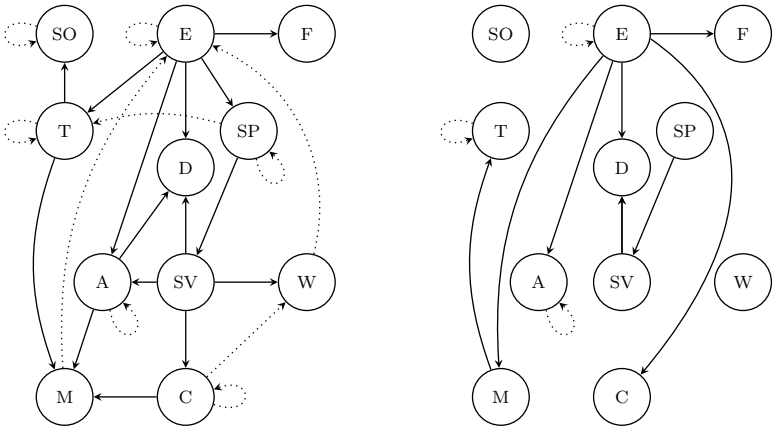


Figure 5.5: Model A: Structure of the best learned network on the Aerial data (left). Majority graph learned from the bootstrap replications of the Aerial data (right). The variables are A = activity; C = cough; D = dyspnea; E = exacerbation; F = FEV_1; M = malaise; SO = SpO_2; SP = sputum purulence; SV = sputum volume; T = temperature; W = wheeze.

as with the synthetic data to ascertain whether this structure is correct. We can however compare it to the structures found by the bootstrap replications. The same procedure as used on the synthetic data leads to a majority graph of the bootstrap replications of the Aerial data, shown on the right in Figure 5.5. Comparing the two graphs we see that the majority bootstrap model is sparser than model A; the arc $E \to C$ is not present in model A and the relations between the variables T, M and E differ in arc direction and temporal signature. All other arcs in the majority bootstrap model are also present in model A. When we test the prediction performance of our models in Section 5.6.3, we will call the model average over the bootstrap models *Model B*.

To get some insight in which variables are good predictors according to the model, it is informative to approach the prediction problem from the other side and look at the probability of predictors given the event, $P(X_{t-\delta} \mid E_t)$. In Table 5.2 the probabilities of some observations for $\delta \in \{0, 1, 2\}$ are shown. We can see that *dyspnea* is a good direct predictor, as its probability increases leading up to an exacerbation. The variable *activity* denotes a decrease in daily activities and is similarly predictive. *Sputum volume* and *cough* give less information about exacerbations over time.

Structural comparison to expert and naive Bayes models

It is informative to compare the result of the structure search to the hand-crafted model introduced in Section 5.5.2. Although this expert model is static, we can at least see whether the found dependencies are similar irrespective of whether arcs are temporal. The comparison is hindered by the presence of hidden variables in the expert model which are not present in the learned models – learning models with hidden variables is a difficult problem, see e.g. [33] for a possible approach. As a consequence we expect to find dependencies between symptoms that are mediated by a hidden variable in the expert model.

Although we cannot really interpret the model in Figure 5.5 causally, it helps to keep the clinical meaning of the variables in mind when comparing the model with the expert model. The dependence between sputum volume and cough is found

X	$P(X_{t-3} \mid E_t)$	$P(X_{t-2} \mid E_t)$	$P(X_{t-1} \mid E_t)$	$P(X_t \mid E_t)$
Dyspnea	0.55	0.60	0.67	0.79
Sputum volume	0.43	0.46	0.49	0.48
Cough	0.50	0.53	0.53	0.50
Activity	0.75	0.78	0.84	0.91

Table 5.2: Probabilities of predictors given evidence of an exacerbation.

in both the expert and learned model. The learning procedure finds a dependence between exacerbation and FEV_1 which is an indicator of lung function, which is in line with the expert model. Fever is a strong indicator of infection, which is often the cause of an exacerbation, explaining the dependency $E \rightarrow T$ in model A, whereas the expert models the dependence of temperature on the hidden variable infection by the arc $I \rightarrow T$. In the expert model activity is influenced by lung function, which is captured by the dependence between activity, exacerbation and sputum volume in model A. So although there are clear differences between the models, model A appears to identify dependencies that can be explained. It should be noted however that identifiability of relations between subjective symptoms can be a problem with limited data, so there might exist other models that are also reasonable and perform similarly. The advantage of Bayesian networks in this context is that they at least provide a simple way to inspect the relations and for example check with a clinician whether the found relations are clinically defensible.

If we compare the structure of model A to the temporal naive Bayes structure (not shown), we see that most of the dependencies between the features remain intact. Note that the structure of this model is restricted by both the naive Bayes arcs from exacerbation to all the other variables and by the complexity limit in our search procedure which bounds the number of parents to three. For the augmented naive Bayes model five temporal arcs are found, self loops for E, A, SP and T and the arc $SP \rightarrow T$, all of which are also present in model A. The atemporal arc $T \rightarrow SV$ is not present in model A and the path $A \rightarrow SO \rightarrow D$ is represented by the arc $A \rightarrow D$ in model A. Based on the data we have it is hard to determine directly which of these models is correct. Therefore we now turn to the evaluation of the models in terms of exacerbation prediction performance.

5.6.3 *Prediction performance & validation*

Ultimately, we are interested in prediction performance. We can use our bootstrap replications as an alternative cross-validation, measuring performance of model A on the bootstrap data. To do so we concatenated all the bootstrapped data and computed the TPR, FPR and AUC for the combined data. The results with model A are modest, with TPR = 0.76, FPR = 0.40 and AUC = 0.66. Although ideally our model should be tested on prospective data – a project to gather validation data is in planning – we can gain some insight in the generalisability of our results by testing the performance on a different data set.

We performed an ROC-analysis on the validation data set D_{val}, that has not been used for model learning. The ROC-curves for model A and B are shown in Figure 5.6. Model A reaches an AUC of 0.84 with one day ahead predictions on the validation data, which is an encouraging result. Model B, the model average over the bootstraps, outperforms model A, obtaining an AUC of 0.90. In practice one has to decide on a cut-off point somewhere on the ROC-curve which gives a reasonable trade off between true and false positive rates. Our system can adapt the kind of feedback that is given to the patient based on the probability of exacerbation, incorporating

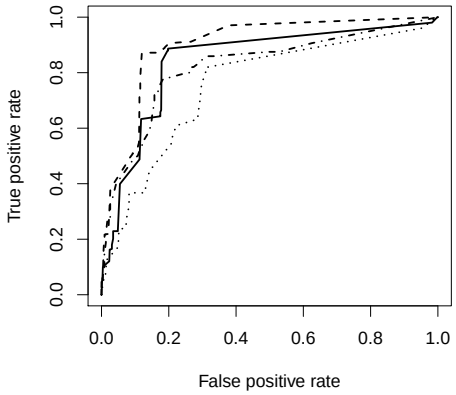


Figure 5.6: ROC-curve on the validation data set D_{val} for the Aerial model predictions (solid); for the bootstrap averaged model (dash); and on the deduplicated validation data D_{dedup} for the Aerial model (dot) and bootstrap averaged model (dash-dot).

different cut-off points for different kinds of advice. We should be careful however in interpreting the prediction results as consecutive time points are often the same, positively skewing the results by predicting that nothing changes.

To test the influence of repetitions in the test data we also checked the performance of the models on D_{dedup}, the deduplicated version of the data. This way we can see how the model reacts to state changes, which is ultimately what we are trying to detect. In Table 5.3, we summarise the results for different models: the static and dynamic expert model, the augmented temporal naive Bayes model and model A and B on the regular validation data and the deduplicated data. For all models we compute the performance measures from the prediction probabilities $P(E_{t+\delta} \mid O_t)$ for $\delta \in \{1, 2\}$.

As expected performance drops for all models on D_{dedup}, with the most notable decrease in the expert models. As AUC alone is not sufficiently informative, we also look at true and false positive rates, computed with the point on the ROC-curve closest to $(0, 1)$ as cut-off point. The true and false positive rates show that although the baseline static model appears to perform on par with the temporal models, this is only true on average over cut-off points. At the theoretically optimal cut-off we see that detecting changes is difficult with a static model, as one would expect. Model A and the temporal naive Bayes perform similarly in this condition, but are outperformed by model B in terms of false positive rate without a significant change in true positive rate. The Brier scores show a similar pattern, although differences in scores are small.

The performance for two day ahead predictions is also shown in Table 5.3. A decrease in performance is expected as health status can change quickly over time. The same caveat with respect to predicting 'no change' applies, so again the performance on the deduplicated data is of more interest. In terms of AUC the performance drop seems relatively small, however false positive rates are getting less acceptable for a useful application. The static expert model shows some remarkable behaviour by increasing its TPR for two day ahead predictions, but this is mostly a consequence of

	Model																			
Data	Expert static				Expert dynamic				Temporal ANB				Model A				Model B			
$t+1$	AUC	TPR	FPR	BS	AUC	TPR	FPR	BS	AUC	TPR	FPR	BS	AUC	TPR	FPR	BS	AUC	TPR	FPR	BS
D_{val}	0.86	0.89	0.25	0.11	0.85	0.89	0.25	0.19	0.89	0.90	0.21	0.09	0.84	0.89	0.20	0.10	0.90	0.87	0.12	0.09
D_{dedup}	0.76	0.59	0.14	0.14	0.76	0.60	0.15	0.20	0.77	0.80	0.31	0.15	0.75	0.82	0.32	0.15	0.82	0.78	0.19	0.13
$t+2$																				
D_{val}	0.84	0.86	0.26	0.11	0.83	0.86	0.26	0.19	0.87	0.87	0.22	0.10	0.86	0.82	0.15	0.10	0.86	0.84	0.13	0.11
D_{dedup}	0.72	0.72	0.39	0.15	0.72	0.54	0.16	0.20	0.72	0.78	0.36	0.16	0.73	0.70	0.31	0.15	0.76	0.69	0.21	0.15

Table 5.3: Summary of AUC, True Positive Rate, False Positive Rate and Brier Score for different models on the validation data set D_{val} and the deduplicated data D_{dedup}.

finding the optimal cut-off, resulting in a higher TPR but also very high FPR. The naive Bayes model performs best in terms of TPR, but model B remains the best performer in terms of FPR. Hence it appears that different approaches are all capable of making predictions with some accuracy, but that model averaging over bootstrap models has a slight edge. This shows that bootstrapping is a useful technique to learn clinical models from small data samples of patient monitoring data.

5.7 DISCUSSION

Chronic disease management using automatic data interpretation requires analysing time series to make predictions. We studied time series data from a pilot study with chronic obstructive pulmonary disease patients, with the purpose of developing a predictive model for COPD exacerbations. In this section we discuss the impact of our findings, limitations and future work.

There has been quite some work on telehealth and monitoring for chronic diseases in general and COPD in particular [67], but automatic data interpretation has not played an important role up to now. The methods we employed based on developments in artificial intelligence however, offer powerful tools for automatic interpretation. Bayesian networks are well suited to deal with data analysis in a clinical context, because the models are interpretable – one can ask a clinician whether the relations found with structure learning make clinical sense. Equally important for use in a home-care setting is that it remains possible to make predictions when data values are missing. Since missing values are virtually impossible to prevent completely, this is an important feature. Dynamic Bayesian networks generalise these desirable features to time series analysis. An issue that relates to model interpretability is that the best performance is obtained by a model average, which is more difficult to understand. Cooperation with a clinician seems advisable, to make sure that differences in models describe possible domain features. Alternatively, augmented naive Bayes may be a better choice when a fully interpretable model is desired. In the augmented naive Bayes model only the interpretation of the dependencies has to be verified to be clinically valid and at least in our data the performance difference with the model average was fairly small.

Structure learning of Bayesian networks has been studied extensively [57], but the present challenge was combining structure learning techniques for dynamic Bayesian networks with sparse data methods. Structural EM is in itself already a computationally expensive operation and bootstrapping data sets makes it even more daunting, which is clearly a limitation when using these methods for very large models. The computation time for our COPD data however, although long in absolute terms (about a day per data set on current hardware), is negligible compared to the time investment of data acquisition. Therefore it may still be worth considering even

for larger models. The current result also indicate that if model averaging is too expensive, augmented temporal naive Bayes provides an alternative that is easier to construct because the model space is restricted, but offers good performance relative to the computational complexity.

For our COPD application our results are promising as we are able to learn models that appear quite capable of predicting exacerbation occurrences. Nonetheless there are limitations to the extend to which we can rely on the results of the validation. The test data has not been gathered specifically for this purpose and does not contain all the variables in our model, also the number of missing values may not be typical for the data that needs to be analysed in the final system. Furthermore, the decision on what is an acceptable trade-off between true and false positives is beyond the technical scope and depends on clinical views, cost considerations and regulatory requirements. As the project's initial focus was on analysing the monitoring data, background information on the patient, including current treatment, and more general environmental factors (e.g. time of year, weather conditions) have not been taken into account at present, but are part of future work. Whether our results generalise will be tested in a follow-up study that is currently being planned, in which a more substantive group of patients will be monitored, extensive baseline information will be taken into account and exacerbation events will be verified by a pulmonologist as gold standard. A confirmation of the present prediction results would then open up the way for a practical implementation of automatic data interpretation for COPD disease management.

6

PROBABILISTIC REASONING WITH TEMPORAL INDETERMINACY

6.1 INTRODUCTION

In clinical medicine, as in many other fields, one frequently has to deal with data that is uncertain and records temporal progression. The time aspect is often also uncertain, which is sometimes referred to as *temporal indeterminacy* [22]. To build useful predictive models we need to deal with these two kinds of uncertainty. A typical example in clinical medicine would be diagnosis based on a patient report of symptoms, e.g. over the last week. From a modelling perspective it may be uncertain *which* symptoms occurred, and, while our knowledge of *when* the symptoms occurred is constrained to last week, within that period we may not be able to be more specific.

There are many methods that support modelling processes of time; dynamic Bayesian networks (of which hidden Markov models are a special case) are an example [28, 69]. The problem of temporal indeterminacy and its relation to temporal granularities has been recognised in the database community (e.g. [22]). However, the fundamental problem of dealing with temporal indeterminacy in predictive models is mostly unsolved.

The representational power of Bayesian networks makes them a useful tool in a clinical context to represent (causal) relations between symptoms, signs and diseases. We have developed a decision support system for chronic obstructive pulmonary disease (COPD), where patients are monitored at home to detect exacerbation events. In this system a Bayesian network is used for automatic data interpretation [96]. The dynamic nature of the disease process leads naturally to the desire to employ a model that mirrors the disease and clinical practice more closely than a static Bayesian network. The first step is then to extend the Bayesian network to a dynamic Bayesian network (DBN) which allows us to model dependences through time.

Temporal indeterminacy results from the harsh reality of patient monitoring in a home environment, in the sense that it may not always be possible to obtain precise information at the right time. We want to be capable of dealing with indeterminacy of the kind "symptom S appeared between 2 and 4 days ago". This problem is not limited to monitoring, but appears often in clinical practice. We can identify similar aggregation problems: when lots of data is available (e.g. on the intensive care unit), aggregation as a summary is sometimes useful; a situation that occurs frequently is that measurements are taken at different times but the clinician is interested in an aggregated result that represents the health status of the patient over the period the measurements were taken. These problems have been studied in the context of temporal abstraction [87], but not in probabilistic graphical models.

We aim to develop a framework to specify temporal indeterminacy, and related aggregation problems, probabilistically. In this paper we focus on this representational problem in the context of dynamic Bayesian networks, with the objective to make predictive models that handle indeterminacy.

6.2 PRELIMINARIES

6.2.1 *Representing time*

In order to reason about temporal indeterminacy we need a formalism to represent time and granularity. The work on representing time in temporal databases seems useful as a starting point [22]. We take a linearly ordered set of points $(\mathcal{T}, \leq)$ as the primitive representation of a time line, where $\mathcal{T}$ is a subset of the natural numbers $\mathbb{N}$. A *determined instant* at the lowest time granularity (e.g. seconds) is a point $t \in \mathcal{T}$. Due to temporal indeterminacy and granularities, it may be useful to represent an instant as a set. For example, an instant at the scale of minutes is a single minute, but may be represented as a set of 60 seconds at a finer granularity. Similarly, when the exact time of an instant is unknown, a set of points can be used to represent the uncertainty in time, that is, to represent the temporal indeterminacy. An *indeterminate instant* is defined as a set $a \subseteq \mathcal{T}$, with the property that a is contiguous: $\forall t \in \mathcal{T} : \inf a \leq t \leq \sup a \Leftrightarrow t \in a$, using $\inf a$ and $\sup a$ to denote the lower and upper bound of the interval of indeterminacy. Two instants a, b are called non-overlapping if $\sup a < \inf b$ or $\sup b < \inf a$.

In studies on logical representations of granularities (e.g. [11]), different time structures are allowed as granularities. This allows one to specify hierarchies of granularities for example for the Gregorian calendar. Here we define a granularity G in a more restricted sense as a partition of $\mathcal{T}$, i.e. a set of subsets such that $\bigcup G = \mathcal{T}$ and if $g, g' \in G$ then $g \cap g' = \varnothing$ or $g = g'$. We further consider only granularities that are contiguous, uniform and comparable, which means that every time point can be expressed in each granularity and time points within a granularity have the same size (which excludes for example 'month'). Examples of possible granularities with these restrictions include 'second', 'hour', 'day' etc.

Before we go on we introduce some notation. We are often interested in events associated with time. In particular we consider events of observing a value of a certain variable of interest. We write E_a for an event E that occurs in a, where a is an indeterminate instant. It is sometimes convenient to specify a certain time point t within a. For example, to denote the probability that E occurs at $t \in a$, with a an instant, we write $P(E_a(t))$. The complement event of E_a is denoted $\bar{E}_a$.

6.2.2 *Bayesian networks*

To be able to add temporal indeterminacy to temporal Bayesian networks, we first define the usual way to explicitly incorporate time in Bayesian networks.

A *dynamic Bayesian network* is a pair $(G, \mathbf{F})$, with G a graph $G = (\mathbf{V}, \mathbf{A})$ and $\mathbf{F}$ a set of factors over random variables $\mathbf{X}$ corresponding to the vertices $\mathbf{V}$. The set of arcs in the graph $\mathbf{A} \subseteq \mathbf{V} \times \mathbf{V}$ represents (in)dependences of the variables. For a two-slice dynamic Bayesian network we subdivide the vertices in an initial slice and a repeated transition slice $\mathbf{V} = \mathbf{V}_0 \cup \mathbf{V}_{1:T}$, which implies a similar subdivision $\mathbf{X} = \mathbf{X}_0 \cup \mathbf{X}_{1:T}$. An arc $(v, w) \in \mathbf{A}$ with $v \in \mathbf{V}_t$ and $w \in \mathbf{V}_{t'}$, $t < t'$, denotes a temporal dependence. Let $\mathrm{pa}(X)$ denote the parents of $X \in \mathbf{X}$ in the graph G. The set $\mathbf{F}$ contains factors for each $X \in \mathbf{X}$ such that $f(x, \mathbf{p}) = P(X = x \mid \mathrm{pa}(X) = \mathbf{p})$. The joint distribution over $\mathbf{X} = \bigcup_i^T \mathbf{X}_i$, factorises over the graph such that $P(\mathbf{X}) = \prod_{X \in \mathbf{X}} P(X \mid \mathrm{pa}(X))$.

To simplify the models, often only first-order Markov models are considered, which means that the future is independent of the past given the present:

$$P(X_{t+1} \mid X_t, X_{1:t-1}) = P(X_{t+1} \mid X_t).$$

Furthermore, a usual assumption is homogeneity, also called stationarity, which means that the probabilities are equal between time steps:

$$P(X_t \mid \mathrm{pa}(X_t)) = P(X_{t'} \mid \mathrm{pa}(X_{t'})) \quad \forall t, t' : X_t \in \mathbf{X}_t, X_{t'} \in \mathbf{X}_{t'}.$$

This may be too strong an assumption for real processes, which has lead to recent work on learning non-stationary dynamic Bayesian networks [81, 42]. We leave these assumptions intact here, and focus only on the indeterminacy problem.

A technique often used to model the interaction between multiple variables (to prevent an exponential growth of parameters) are *causal independence* models [75]. The main assumption is that different causes of an effect can be assumed to be independent and only interact through a particular deterministic function. Here we will use this method to model the interaction between random variables at different time granularities.

6.3 EVENTS AND GRANULARITY

We first state some properties of indeterminate events, and their consequences. Consider an event E_a to be a point-like occurrence at the finest granularity under consideration somewhere at a. We then have:

$$E_a = \bigvee_{t \in a} E_a(t), \tag{6.1}$$

i.e., event E_a is defined in terms of events at the finer granularity. Analogously, the complement event is:

$$\bar{E}_a = \neg \bigvee_{t \in a} E_a(t) = \bigwedge_{t \in a} \bar{E}_a(t). \tag{6.2}$$

The following property ensures that only a single point event $E_a(t)$ is true:

$$\forall t, t' \in a : t \neq t' \rightarrow E_a(t) \wedge E_a(t') = \bot. \tag{6.3}$$

called the *single event assumption*. This means that the event at coarser granularity is caused by exactly one event at finer granularity. Conversely, if multiple events could have occurred at the finer granularity, this is called the *multiple event assumption*.

When using probability theory to capture the principles of an indeterminate instant a, we need to specify a distribution for the occurrence of an event E associated with the instant a. Under the single event assumption a distribution for E_a has the property:

$$P(E_a) = \sum_{t \in a} P(E_a(t)),$$

which follows from the two properties above. From Property (6.3) it also follows directly that:

$$\forall t, t' \in a : t \neq t' \rightarrow P(\dots, E_a(t), \dots, E_a(t'), \dots) = 0 \tag{6.4}$$

and

$$P\Big(\bigwedge_{t \in a \setminus \{t'\}} \bar{E}_a(t) \mid E_a(t') \Big) = 1.$$

From this, it follows that

$$\begin{aligned} & P\Big(\bigwedge_{t\in a\setminus\{t'\}} \bar{E}_a(t), E_a(t')\Big) \\ & = P\Big(\bigwedge_{t\in a\setminus\{t'\}} \bar{E}_a(t) \mid E_a(t')\Big) P(E_a(t')) \\ & = 1 \cdot P(E_a(t')) = P(E_a(t')). \end{aligned} \tag{6.5}$$

However, when the multiple event assumption is adopted, only

$$P(E_a) = P\left(\bigvee_{t\in a} E_a(t)\right)$$

holds. This means that at the aggregated level we cannot distinguish between what are multiple events at the precise granularity. We are interested in somehow relating the representations at both granularities.

Further constraints result from domain specific patterns of dependence between points in a, which will be considered later.

6.4 TEMPORAL INDETERMINACY IN DBNS

Now we focus on modelling temporal indeterminacy in dynamic Bayesian networks. In general the problem is representing a temporal process at multiple granularities.

6.4.1 *Probabilistic aggregation*

A conceptually simple representation consists of separate models for the granularities. This is shown in Figure 6.1. Basically, the lower part of the model aggregates the top part, although here it is not specified how the aggregation takes place. A one-to-one translation from the possible states of the finer granularity to the coarser granularity results in:

$$P(O_a \mid X_a)P(X_a) = P(\bigwedge_{i=1}^{n} O_i \mid \bigwedge_{j=1}^{n} X_j)P(\bigwedge_{k=1}^{n} X_k),$$

where basically the domain of O_a is the Cartesian product of the domains of $O_1, \ldots, O_n$; similarly, the domain of X_a is the Cartesian product of the domains of $X_1, \ldots, X_n$. The result is a representation that is exponential in the size of the domains.

With the single event assumption it is easily possible to compute the aggregated probabilities. However, when represented graphically, these logical constraints would require dropping the first-order Markov assumption yielding a factorisation of $P(X_1, \ldots, X_n)$ without any independence information. However, in the transformation it is not necessary to use a graphical representation as there is a very specific relation between the granularities. Let X_a take on values in $\{0, 1, \ldots, n\}$ then $P(X_a = i) = P(X_i = 1)$ and $P(X_a = 0) = 1 - \sum_{i=1}^{n} P(X_a = i)$, because by definition for i with $1 \leq i \leq n$:

$$\begin{aligned} P(X_a = i) &= P(X_1 = 0, \ldots, X_i = 1, \ldots, X_n = 0) \\ &= P(X_n = 0, \ldots, X_{i+1} = 0 \mid X_i = 1, \ldots, X_1 = 0) \\ &\quad \cdot P(X_i = 1, X_{i-1} = 0, \ldots, X_1 = 0) \\ &= 1 \cdot P(X_i = 1, X_{i-1} = 0, \ldots, X_1 = 0) \\ &= P(X_i = 1) \end{aligned}$$

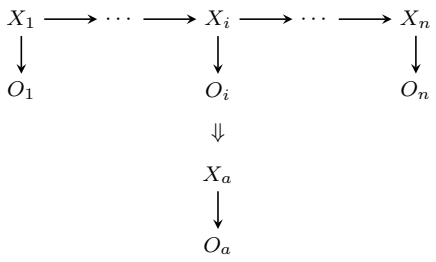


Figure 6.1: A general depiction of a model at two granularities (without the single event assumption).

using Equation (6.5). Note that it follows that

$$P(X_a \neq i) = P(X_i = 0).$$

$$\begin{aligned} P(X_i = 1) &= \sum_{X_j, i \neq j} P(X_i = 1, \bigwedge\nolimits_j X_j) \\ &= P(X_i = 1, \bigwedge\nolimits_j X_j = 0), \end{aligned}$$

where the second equality follows from mutual exclusivity, as is easily seen from Equation (6.4). Since this is true for X_i with $1 \leq i \leq n$ and $P(X_a)$ is normalised by setting $P(X_a = 0) = 1 - \sum_{i=1}^{n} P(X_a = i)$, we obtain the aggregation $P(X_a) = P(\bigwedge_{i=1}^{n} X_i)$.

To aggregate the observation variables O_i we take the domain of O_a to be binary. Since we assumed stationarity we have only two parameters: for $1 \leq i \leq n$ $P(O_i = 1 \mid X_i = 1) = p$ and $P(O_i = 1 \mid X_i = 0) = q$. Still assuming that the first-order Markov assumption does not hold, conditioned on the X_i's the O_i variables are conditionally independent, allowing us to aggregate the observations by multiplying the conditional probabilities; for $1 \leq i \leq n$, it holds that:

$$P(O_a = 1 \mid X_a = i) = P(O_i = 1 \mid X_i = 1) \cdot \prod_{j \neq i} P(O_j = 1 \mid X_j = 0) = pq^{n-1}$$

and in addition

$$P(O_a = 1 \mid X_a = 0) = \prod_{j} P(O_j = 1 \mid X_j = 0) = q^n.$$

6.4.2 *Structural aggregation*

A different perspective on modelling indeterminacy is including the aggregate temporal structure within the model. This turns out to be useful as it allows us to model explicitly the interaction between the variables at different granularities. The mechanism that we will employ to model multiple interactions in a systematic way is *causal independence*. This is depicted in Figure 6.2. A causal independence model allows us to model the projection by means of a particular deterministic function of the variables at the finer granularity.

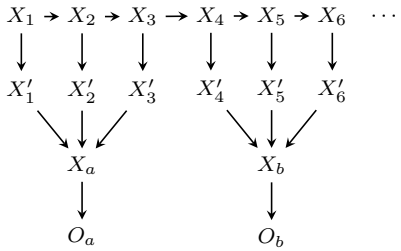


Figure 6.2: Incorporating granularity within the model. Observation variables for the finer granularity have been omitted.

Although separating the state and observation model is common practice to model measurement errors, we will for now assume that state variables are directly observable in order to focus on modelling the indeterminacy. Hence, disregarding the O variables in Figure 6.2 we obtain the following joint distribution for the aggregation over a

$$P(X_a, \bigwedge_{i=1}^{n} X'_i, X_i) = P(X_a \mid \bigwedge_{i=1}^{n} X'_i) \prod_j P(X'_j \mid X_j) \\ P(X_1) \cdots P(X_i \mid X_{i-1}) \cdots P(X_n \mid X_{n-1}). \tag{6.6}$$

In a causal independence model the term $P(X_a \mid \bigwedge_{i=1}^{n} X'_i)$ is represented by a deterministic function. Equation (6.1) implies that the logical OR is a natural interaction function.

As the single event assumption is easily represented by the construction in the previous section, we now focus on the multiple event assumption. The causal independence model allows us to combine the events, even when the independence assumption on the causes is not met (as can be seen to be the case in Figure 6.2). Although the model complexity increases, the number of parameters is still linear in the number of variables at the fine granularity, as opposed to the exponential increase which results when connecting the granularities directly.

The model depicted in Figure 6.2 is not the only possible aggregation. An aggregation variable can summarise arbitrary sets of state variables, which is related to the work by [34]. The statement "symptom S occurred between 2 and 4 days ago" can be modelled with an aggregation variable that summarises 3 variables at the granularity of days. These aggregations can be made to overlap, that is X_a summarises $X_{1:3}$, X_b summarises $X_{2:4}$ etcetera.

Aggregation patterns

Starting from the aggregated level we can use domain knowledge to specify a pattern on the finer granularity. It seems worthwhile to identify patterns that have intuitive appeal, defining which variables have more influence on the aggregated variable, or, from the perspective of the coarser granularity, which variables are more likely to have caused the aggregated state. Specifically, we consider four situations: the *uniformity* assumption implies that we have no reason to believe that there is a particular structure at the precise granularity, the maximum entropy choice. A *discretised normal* distribution is useful to model the uncertainty around a particular observation point. The *decreasing* and *increasing* patterns convey the idea that some event is more likely to occur at the start or the end of the aggregated interval.

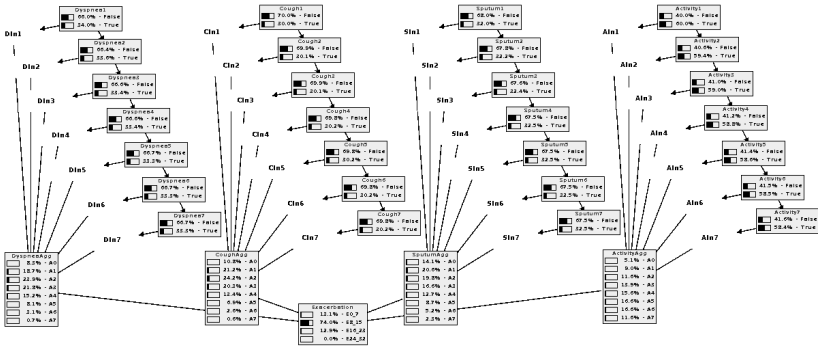


Figure 6.3: COPD-monitoring aggregation network.

If a point in the coarser granularity X_a has corresponding points in the finer granularity $X_1, \ldots, X_i, \ldots, X_n$, the patterns have the following properties:

Uniform: $P(X_1) = P(X_i) = P(X_n)$

Normal: $P(X_i) = P(X_{n-i+1}) < P(X_{n/2})$

Decreasing: $P(X_1) \geq P(X_i) \geq P(X_n)$

Increasing: $P(X_1) \leq P(X_i) \leq P(X_n)$

Note that for the *normal*-pattern some care needs to be taken when aggregating an even number of variables. For the *increasing* and *decreasing* patterns one could also consider the strict cases, but especially for large n the current versions are more flexible.

6.5 TEMPORAL INDETERMINACY IN COPD

The model shown in Figure 6.3 is an example of how temporal aggregation can be used in the context of COPD-monitoring. The data has been gathered at various hospitals and general practices in the Netherlands, with the smartphone-based monitoring system as described in [96]. The model is an unrolled dynamic Bayesian network for four symptoms: dyspnea, cough, sputum production and activity capacity. Each variable indicates a daily measurement over the course of a week, and takes on binary values (False = normal, True = worse than normal). The daily symptom variables are aggregated in a variable representing the whole week by means of a generalised causal independence model [90]. Instead of a binary outcome variable, the aggregate has values representing the number of symptom days (0 through 8). The intermediate variables model the inhibition, or noise, parameters. Finally, we are interested in the outcome variable *Exacerbation*, which provides an indication of a clinical significant event. Here again a causal independence model is utilised to aggregate the individual symptoms, by means of a counting function that classifies the number of symptoms into the categories [0-7],[8-15],[16-23],[24-32], that is, it is a deterministic function of the symptom aggregates.

The parameters of the model have been learned from the monitoring data of 8 COPD patients that participated in a pilot study. As before we assumed stationarity,

and with a Dirichlet uninformative prior distribution with $\alpha = 1.1$ we estimated the probabilities of a symptom variable $X \in \{D, C, S, A\}$ as:

$$P(X_i = k \mid \mathrm{pa}(X_i) = j) = \frac{\alpha_{ijk} + N_{ijk} - 1}{\alpha_{ij} + N_{ij} - |X_i|},$$

where N_{ijk} is the number of cases that variable i takes value k and the parent variables of i take on configuration j; α_{ijk} denotes the pseudo-counts; $|X_i|$ is the number of values that variable X_i can take on and $N_{ij} = \sum_k N_{ijk}$. The parameters of the intermediate variables were all equal with $P(XIn_i = 1 \mid X_i = 1) = 0.9$ and $P(XIn_i = 1 \mid X_i = 0) = 0.1$. All the aggregate variables have deterministic parameters as explained above.

In the pilot study, participants were asked to fill in the Clinical COPD Questionnaire (CCQ) as a control measure for the data gathered with the monitoring system. The questionnaire consists of 10 questions on a 7 point scale (lower scores are better); 4 questions are about dyspnea, 1 about cough, 1 about sputum production and 4 about activity; we omitted the 2 psychological questions about dyspnea for the current analysis; the version used asked patients to answer with respect to last week. For our current purposes this gives us a good example to look at aggregation, since we can compare the monitoring data from the preceding week with the results of the CCQ. Because not all data was available at the time of writing and due to missing data (patients not filling in the CCQ), we have data of 5 patients for comparison.

The procedure is as follows, we entered the monitoring data as evidence in the DBN (missing values are easily taken care of by leaving them as non-evidence variables) and noted the probabilities of the aggregate variables. For the CCQ scores we took the mean of the scores for the same aspect, rounded to the nearest integer. The results of the comparison are shown in Table 6.1, where the distribution over the aggregate variables is shown, which basically means the probability of a certain number of days of that particular symptom during the preceding week. Each value of an aggregate variable is mapped to the same CCQ score (0 to 6), except 7 which is mapped to CCQ score 6 (i.e. '(almost) always' is taken to mean 6 or 7 days).

Using the most likely value of the variables as a simple measure to assess agreement with the CCQ scores we see that we end up with the same value in 30% of the cases and are one value removed in 45%. But given the currently limited amount of data it is hard to say whether this could be improved by a better aggregation model or whether there exists a discrepancy between the monitoring data and the CCQ scores. The sputum value of patient 3 favours the latter explanation, as the CCQ indicates sputum was produced almost never whereas the monitoring data provide evidence to the contrary; clearly an aggregation model cannot solve these inconsistencies. With respect to the outcome variable *Exacerbation*, visual inspection of Table 6.1 indicates that patients 1 and 2 are stable, patients 3 and 5 are in the course of an exacerbation and patient 4 is somewhere in between. This shows that the aggregation provides a useful and more easily interpreted summary of the monitoring data.

	0	1	2	3	4	5	6	7
D	47.8	37.2	**12.4**	2.3	0.3	0	0	0
C	0.6	**11.0**	53.8	27.9	6.1	0.7	0	0
S	**4.7**	45.7	35.8	11.6	2.0	0.2	0	0
A	**47.8**	37.2	12.4	2.3	0.3	0	0	0
D	30.6	**37.8**	23.0	7.2	1.3	0.1	0	0
C	32.7	**37.8**	21.7	6.6	1.1	0.1	0	0
S	36.3	36.9	**19.8**	5.9	1.0	0.1	0	0
A	**31.0**	35.3	23.8	8.2	1.5	0.2	0	0
D	0	0.5	4.8	20.7	33.5	**28.5**	11.1	0.9
C	0	0	0.5	4.3	**17.5**	34.0	29.9	13.7
S	0	**0**	0.3	2.4	11.5	28.4	33.7	23.6
A	0	0	0.1	1.2	6.7	**20.8**	36.6	34.7
D	0	1.0	7.8	23.4	**32.9**	25.1	9.0	0.7
C	1.2	13.8	29.7	**32.0**	18.1	4.7	0.5	0
S	33.2	37.1	**21.2**	7.0	1.3	0.1	0	0
A	0	0.2	1.9	**9.6**	24.6	37.8	23.7	2.2
D	0	0	0.2	2.4	13.5	37.8	**41.8**	**4.2**
C	0	1.0	**9.7**	36.8	37.3	13.3	1.8	0.1
S	0	0	0.4	3.9	19.4	44.3	**29.2**	**2.7**
A	0	0	0	0.5	3.7	**15.5**	37.4	42.8

Table 6.1: Distributions (percentage) over the aggregate variables (D=dyspnea, C=cough, S=sputum, A=activity) for 5 patients. In bold the probability of the value that corresponds to the CCQ score.

6.6 RELATED WORK

A well known representation of time is Allen's algebra [3]. It provides a set of relations between time intervals and operations to reason over the intervals. The relations allow expressing information about the order of intervals and some different types of overlap, for example 'X starts or is during Y'. In some sense this allows indeterminacy, but in a qualitative way. Temporal indeterminacy and its relation to representations at different granularities is a topic studied in the context of temporal databases [22]; but the problem is also relevant for planning and scheduling and information systems that deal with temporal data. For medical information systems there has been quite some work on temporal reasoning: on multigranular representations [55]; and on temporal abstraction [87]. The latter deals with aggregating temporal data to meaningful higher level concepts, which requires taking care of temporal representation issues like granularities. Our probabilistic graphical model approach provides a different view on a part of these problems. Also related are irregular-time Bayesian networks [80], which generalise DBNs to time-slices with changing size. [11] studied granularities in temporal constraint satisfaction problems; and [19] in a more general linear time logic context, both deal extensively with properties of multiple granularities but not with probabilistic representations.

6.7 CONCLUSION

Modelling temporal indeterminacy in probabilistic graphical models creates the opportunity to deal with different kinds of uncertainty that arises in many practical situations and in chronic disease monitoring in particular. We think that our initial results on probabilistic multigranular models by means of causal independence form a good starting point for further research in this direction. Particularly, the consequences and applicability of the single or multiple event assumption warrant attention. Future work also includes further analysis in the context of COPD-monitoring.

7

DESCRIBING DISEASE PROCESSES USING A PROBABILISTIC LOGIC OF QUALITATIVE TIME

7.1 INTRODUCTION

In solving clinical problems such as diagnosis or prognosis, concerning the signs and symptoms of a disease, one often has to take into account the time when a particular event has occurred or is expected to occur. In many cases, the actual temporal details about when events have occurred are not available, or at least imprecise, whereas one is more certain about the order of the events. AI researchers have traditionally used *Allen's interval algebra* [3] to model imprecise temporal events. It forms the foundation of a temporal logic that supports reasoning about temporal events in a qualitative fashion [4]. Work by Shahar [87] indicates the usefulness of Allen's algebra for describing temporal events in medicine. However, Allen's algebra does not allow expressing uncertainty about the occurrence of the events or their qualitative, temporal relationships. Yet, uncertainty is a feature of many problems where precise temporal information is missing, such as in clinical medicine.

In the work described in this paper it is investigated in what way Allen's interval algebra can be extended to incorporate uncertainty reasoning. Recently developed probabilistic logics can be used as a basis for such a more general language and in particular we build upon the work on CP-logic [103]. The aim is to design a framework that allows describing disease processes in a way similar to what is found in the clinical literature, i.e. imprecise yet with uncertainty made explicit.

Developing sophisticated decision-support systems for realistic clinical problems requires one to handle both the imprecision and uncertainty of medical knowledge and data. The methods developed in this paper are expected to contribute to meeting these challenges. Evidence that this is justified comes from examples taken from important clinical problems, one of which is the management of chronic obstructive pulmonary disease for which we have developed a smartphone-based decision-support system [97].

We claim that to model disease progression it is often easier to start from imprecise notions of temporal ordering compared to directly constructing for example a dynamic Bayesian network. That is, the kind of information readily available from questioning a patient and from general clinical knowledge about the disease, offers a good starting point. From there on we can add uncertainty to obtain a temporal probabilistic model. This way of modelling clarifies the structure of the process and makes it more explicit what kind of assumptions are made.

The topic of temporal reasoning has already received much attention, also in a medical context, exemplified by the book of Combi et al. [20]. However the work described there is almost completely symbolic in nature, focussing on logic properties. We argue that in a realistic clinical setting one cannot and should not ignore the uncertainties that arise due to hidden complexity, lack of information or measurement error. Whereas pure probabilistic frameworks tackle only the uncertainty,

our qualitative time probabilistic logic seeks to model temporal uncertain processes at a practical level of abstraction.

This paper is organised as follows. In the next section we introduce two motivating examples, typical for the problems encountered in biomedicine, which will be used later to validate the framework being developed. In Section 7.3 we discuss related work and in Section 7.4 we provide some preliminaries. In Section 7.5 we describe the probabilistic temporal framework that seems suitable for the description of disease processes, and give some properties of the language. In Section 7.6 we return to our examples and show that the developed framework can be usefully applied to describe the temporal, uncertain evolution of disease processes.

7.2 MOTIVATION

7.2.1 *HIV drug resistance*

Our first example, from [48], is modelling HIV mutations and drug resistance. In contrast to the COPD example which will be introduced next, HIV mutations are uncertain temporal events without recurrence. In [48] they represent the problem using temporal nodes Bayesian networks (TNBN), where random variables take time intervals as values to denote time frames in which a mutation could occur. In Section 7.6 we show that TNBNs can be represented in our framework.

7.2.2 *The management of COPD*

Throughout the paper we will use the management of chronic obstructive pulmonary disease, COPD for short, to motivate the developed methods. COPD is a progressive lung disease characterised by a mixture of chronic bronchitis and emphysema, leading to decreased respiratory capacity and potentially to respiratory failure and death. Although there are a number of causes, exposure to (tobacco) smoke is the most prevalent. Because COPD is a progressive disease, its temporal development is quite important and even more so because of the occurrence of *exacerbation* events – a worsening of symptoms with possibly a large negative influence on health status.

To model exacerbations we have to take into account that patients are usually in a home care situation which lacks precise high-frequency measurement-equipment. As a consequence we have to rely on imprecise information. Probability theory provides a tool to quantify our uncertainty about the state of the system, in this case the health status of a COPD patient. We focus on a limited number of random variables that characterise the events of the uncertain process: two main symptoms, *dyspnea* and *cough*, a common cause *infection* and the outcome *exacerbation*. As we will be using a probabilistic logic the relations between these variables can be specified in terms of causal rules:

$$\mathsf{dyspnea} \leftarrow \mathsf{infection}.$$
$$\mathsf{cough} \leftarrow \mathsf{infection}.$$
$$\mathsf{exacerbation} \leftarrow \mathsf{dyspnea} \wedge \mathsf{cough}.$$

Limited information also leads to temporal uncertainty, yet it is often possible to determine qualitative temporal relations. We use time intervals to model the duration of symptoms and other variables of interest. Allen's algebra then provides

a language to state ordering relations between symptoms in time. The rules above can be extended to incorporate the temporal information:

$$\mathsf{dyspnea}(I) \leftarrow \mathsf{infection}(J), \mathsf{allen}(I, J).$$
$$\mathsf{cough}(K) \leftarrow \mathsf{infection}(J), \mathsf{allen}(K, J).$$
$$\mathsf{exacerbation}(L) \leftarrow \mathsf{dyspnea}(I) \wedge \mathsf{cough}(K) \wedge \mathsf{allen}(L, I) \wedge \mathsf{allen}(L, K).$$

where I, J, K, L denote the intervals which we associate with the events and allen denotes some qualitative time relation between the relevant intervals. To predict exacerbations, we have to quantify the uncertainty of temporal relations between symptoms. This will be the main example, and in Section 7.6 we will study the situation introduced here in detail.

7.3 RELATED WORK

Allen's algebra has received much attention over the years and finds applications in fields from planning to clinical medicine and many more. However there are also numerous other temporal representation and reasoning frameworks. A comprehensive overview lies outside the scope of a related work section, but we mention some important work. For an overview of topics related to time in medicine we refer the reader to [20].

Besides Allen's work [3, 4], McDermott's work on temporal logic [65] is well-known. Interesting to note is McDermott's observation that quite a few problems result as a consequence of uncertainty, and that no formal framework exists that satisfactorily combines logic and probability. Fortunately this has changed in recent years, leading to our current work on temporal reasoning in probabilistic logic. Combi et al. [21] describe an extension of Allen's logic that generalises to different temporal granularities, which is often necessary in a clinical context.

Also important to mention is the work on temporal constraint networks by Dechter, Meiri and Pearl [29] and on combined metric and Allen constraints by Kautz and Ladkin [53]. Jonsson and Bäckström [52] later showed that disjunctive linear relations subsume these and a number of other temporal constraint formulations. An application of temporal constraints in a medical context can be found in modelling clinical guidelines, see e.g. [92].

In the probabilistic model field there has also been interest in temporal models, and although not directly connected to qualitative time representations, dynamic Bayesian networks [28, 69] are a well-known instance of temporal probabilistic models. Many variants and extensions exist such as the networks of probabilistic events [39] where nodes are associated with events and the value of a node represents an occurrence of the event at a particular time point; temporal Bayesian networks of events [7] which are similar and use a subset of Allen relations to relate (quantified) temporal nodes; and the work by Tawfik and Neufeld [91] on temporal reasoning in Bayesian networks. These frameworks do not focus on representing qualitative time as we do in this paper.

Other work on combining probabilistic and qualitative temporal reasoning includes probabilistic temporal interval networks [82], a probabilistic extension of the interval constraints network often used with Allen's algebra. This is similar to our current proposal, but our work also allows modelling uncertainty in the events. Probabilistic temporal networks, as defined by Santos and Young [83], are network models that incorporate Allen's constraints on conditional probabilities. They make the assumption however that the intervals of interest are known beforehand and can be specified explicitly, therefore not allowing uncertainty in what intervals are or will be of interest.

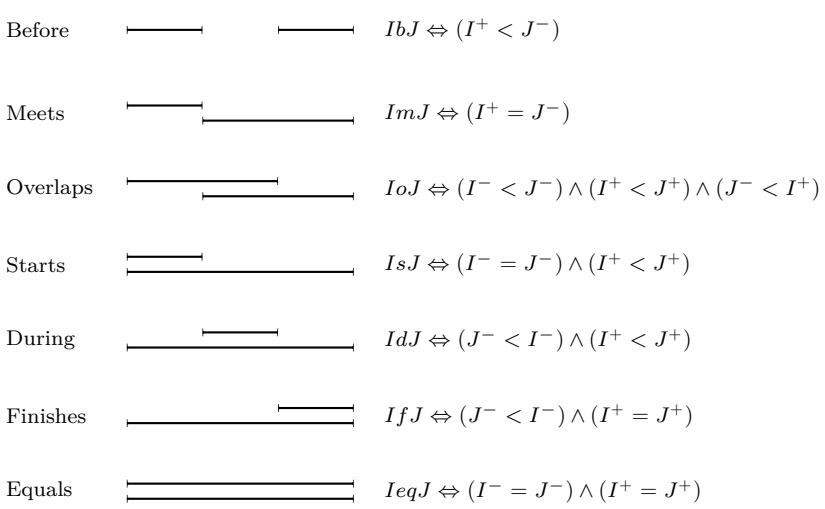


Figure 7.1: Graphical representation of the seven basic relations that can hold between two time intervals. These relations and their inverse make up Allen's algebra.

Then, recently, probabilistic logic has been applied to represent stochastic processes [93]. The proposed language CPT-L extends CP-logic to represent fully-observable homogeneous Markov processes. Our work differs in focussing on using CP-logic for qualitative time stochastic processes. Distributional clauses [43] extend the probabilistic logic approach (based on Sato's distribution semantics [84] like CP-logic) to continuous distributions.

7.4 PRELIMINARIES

7.4.1 *Allen's interval algebra*

Allen's algebra builds upon qualitative relations between time intervals. An interval implicitly refers to an event that takes place during that interval. In Figure 7.1 the relations that can hold between two intervals are shown graphically. We define intervals as right-open $I = [I^-, I^+)$ on a linearly ordered time line of points $(\mathcal{T}, \leq)$, with $\mathcal{T}$ a subset of the set of real numbers $\mathbb{R}$ or, if necessary, restricted to a finite subset of the natural numbers $\mathbb{N}$. The special points I^-, I^+ are defined as $I^- = \inf I$ and $I^+ = \sup I$.

We can now define relations, letting $\mathcal{I}$ be the set of all intervals of $\mathcal{T}$. A binary *temporal interval relation* R is defined as $R \subseteq \mathcal{I} \times \mathcal{I}$. Instead of temporal interval *relations*, we shall in the following use a predicate logic representation, where the logical atom $R(I, J)$, having the meaning of $(I, J) \in R$ in the relational form, will be denoted in infix form as IRJ.

Allen defined a set of seven basic interval relations on two time intervals. Together with the inverses of these seven relations we obtain a minimal set of relations that can express any qualitative relation between two intervals. This set of relations with respect to intervals in $\mathcal{I}$ will be denoted $\mathcal{B}$ and the thirteen relations therein are $\mathcal{B} = \{b, \bar{b}, m, \bar{m}, o, \bar{o}, s, \bar{s}, d, \bar{d}, f, \bar{f}, eq\}$, where $\bar{r}$ is the inverse relation of r, which for intervals I, J is defined as $I\bar{r}J \equiv JrI$. Figure 7.1 gives the definition of the relations in terms of interval endpoints, with I^- denoting the start point and I^+ the end

point of interval I, and similarly for J. The basic relations in the set $\mathcal{B}$ are mutually exclusive and collectively exhaustive with respect to the possible relations between two intervals.

In the following examples we will consider events E with an interval index I, written as E_I, to denote that event E occurs in interval I, instead of pure interval expressions, as this makes it easier to convey ideas on how to put time intervals to use. We will also use the shorthand notation $E_I\ R\ E'_J$ for the formally correct notation $E_I \wedge E'_J \wedge (IRJ)$ meaning that event E occurs at time interval I and E' at interval J and that the relationship between these intervals is expressed by the Allen expression IRJ.

Example 7.1. *A certain group of COPD patients tends to have relatively frequent exacerbations – events of worsening of symptoms – that are usually caused by airway infections. Using the basic temporal relations we can describe that an infection in interval I at least partially precedes the increase in symptoms in interval J. We then obtain the expression:*

$$Inf_I\ o\ Sym_J,$$

which means that symptoms can outlast the infection. Since an exacerbation is defined as an increase of the relevant symptoms in the interval we can say for an interval K associated with the exacerbation:

$$Exa_K\ eq\ Sym_J.$$

With the basic relations, the full set of Allen's relations can be constructed. Any of these relations is defined as the disjunction of a subset of the basic relations for the same intervals. Thus, one gets expressions such as $(I\ N\ J) \equiv (I\ R\ J) \vee (I\ R'\ J)$, where N is the new relation, and R, R' are basic relations. However, instead of introducing new names for the resulting new relations, Allen uses set notation for the new relations, giving rise to the following definition:

Definition 7.1. *An* Allen relation *is defined as a disjunction of basic interval relations, represented as a set. The power set of the basic relations contains all Allen relations and is denoted $\mathcal{A} = \wp(\mathcal{B})$. An interval formula is then of the form IRJ with I, J intervals and $R \in \mathcal{A}$.*

Thus, for the example above, $(I\ N\ J)$ would be represented by $(I\ \{R, R'\}\ J)$. Because we will be using Allen's relations as logical relations in what follows, it is useful to notice the effects of Boolean operations on basic relations. The definition above states that Allen's relations are disjunctions of basic relations. From mutual exclusiveness it follows that conjunctions of basic relations are false by definition (at most one relation can hold between any two intervals). For the negation of a basic relation $R \in \mathcal{B}$ we obtain $\neg R = \mathcal{B} \setminus \{R\}$. Note that the negation is thus different from the inverse $\breve{R}$.

Example 7.2. *COPD patients often have what is called ventilation-perfusion inequality – a mismatch between air flow and blood flow through the lung – which may develop during an exacerbation due to increased airway obstruction. When an exacerbation occurs we have a temporal event Vpi_I which is during, finishes or is overlapped by Exa_J. Without any further information the relation between ventilation-perfusion inequality and exacerbation can thus be described by:*

$$Vpi_I\ \{\bar{o}, d, f\}\ Exa_J.$$

Given a set of interval relations, other temporal relations can be derived, using the transitivity of time ordering. For example if IbK and KbJ, then by transitivity it holds that IbJ. Similar rules can be constructed for all pairs of basic relations. In many cases the result will not be a basic relation however, but a set of possible relations, that is an element from $\mathcal{A}$. Allen [3] provides a table of all the inference rules.

More in general we can say that given a finite set of interval relations, the relations that are entailed by this set can be derived by repeated application of the following closure operator until a fixed point is reached [70]:

Definition 7.2. *Let C be a finite set of interval relations. The* closure operator Γ *maps interval formulas to interval formulas using the operations of inversion ($\bar{\cdot}$), intersection ($\cdot \cap \cdot$) and composition ($\cdot \circ \cdot$). $\Gamma(C)$ is the smallest set satisfying:*

1. $C \subseteq \Gamma(C)$
2. *For each I, J that appear in a formula in C : $I\mathcal{B}J \in \Gamma(C)$*
3. *For each $IRJ \in \Gamma(C) : J\bar{R}I \in \Gamma(C)$*
4. *For each $(IRJ), (IR'J) \in \Gamma(C) : I(R \cap R')J \in \Gamma(C)$*
5. *For each $(IRK), (KR'J) \in \Gamma(C) : I(R \circ R')J \in \Gamma(C)$*

The inversion operation is the inverse of each basic relation in R, the intersection operation is simply the set intersection of the relations R, R' and composition is the result of resolving the transitivity of the basic relations. Formally, composition is defined as $\forall I, J : I(R \circ R')J \Leftrightarrow \exists K : IRK \wedge KR'J$.

If there are multiple paths via which a relation between two intervals can be derived we are interested in the strongest relation that holds. So from $\Gamma(C)$ we can derive the *reduced closure* $\Gamma'(C)$, where for each R with $IRJ \in \Gamma(C)$ for given intervals I, J it holds that $IR'J \in \Gamma'(C)$ if $R' \subseteq R$. In other words, the intersection of the relations between two intervals is contained in the reduced closure, because the intersection is the strongest relation that follows from C. See also [70].

Example 7.3. *The temporal relation between the intervals of infection and exacerbation is:*

$$Inf_I \ o \ Exa_J$$

which when combined with the relation,

$$Vpi_K \ \{\bar{o}, d, f\} \ Exa_J$$

can be used to infer the relation between ventilation-perfusion inequality and infection by application of the closure operator:

$$\begin{aligned} C &= \{Inf_I o Exa_J, \ \ Vpi_K\{\bar{o}, d, f\}Exa_J\} \\ \Gamma'(C) &= C \cup \{\, Vpi_K\{\bar{b}, \bar{m}, \bar{o}, d, f\}Inf_I, \ \ Inf_I\{b, m, o, \bar{d}, \bar{f}\}\, Vpi_K, \\ &\qquad Exa_J \bar{o} Inf_I, \ \ Exa_J\{o, \bar{d}, \bar{f}\}\, Vpi_K\} \end{aligned}$$

7.4.2 *Logical reasoning with the interval algebra*

As Allen showed [4], this qualitative algebra is well suited to reason about time in a logic context. Allen's relations are then represented by temporal predicates. The logic we will be using derives from the logic programming tradition of using Horn clauses,

$$H \leftarrow B_1, \ldots, B_n$$

where H is the head of the clause, $B_1, \ldots, B_n$ the body and H and the B_is are logical atoms. Variables are denoted with upper case and are implicitly universally quantified, conjunctions are denoted by commas ',' and a semicolon ';' denotes a disjunction, as in Prolog.

Also instead of using a reified logic approach as Allen does (i.e. using meta-predicates like HOLDS, OCCURS), we opt for the arguably simpler framework of temporal arguments [8] (see also [105] for a discussion on the (dis)advantages of both methods). This means that temporal predicates have a temporal argument specifying the relevant time interval. Note that this implies a typed logic with temporal and atemporal terms, which we will leave implicit as this can always be translated to first order logic at the cost of notational convenience.

A connection between an interval relation representation and predicate logic can be made by introducing a predicate that represents the temporal relation between intervals. Concretely, using a Prolog like syntax, we define the predicate $\mathsf{r}/3$, with the first two arguments representing the intervals and the last argument a basic temporal relation; and the predicate $\mathsf{allen}/3$, with two intervals and a list of basic relations as arguments such that $\mathsf{allen}(I, J, L)$ holds if the disjunction over elements X in list L, $\bigvee_{X \in L} \mathsf{r}(I, J, X)$ holds.

Example 7.4. *Consider again our COPD example which we can now represent somewhat more structured:*

$$\mathsf{exacerbation}(P, J) \leftarrow \mathsf{patient}(P), \mathsf{infection}(P, I), \mathsf{r}(I, J, o).$$
$$\mathsf{vpi}(P, J) \leftarrow \mathsf{patient}(P), \mathsf{exacerbation}(P, I), \mathsf{allen}(I, J, [\bar{o}, d, f]).$$

Here vpi *stands for ventilation-perfusion inequality. As the previous example showed the temporal relation between* infection *and* vpi *can be derived. In this case the result is a disjunction of five possible relations:*

$$\mathsf{vpi}(P, J) \leftarrow \mathsf{patient}(P), \mathsf{infection}(P, I), \mathsf{allen}(I, J, [\bar{b}, \bar{m}, \bar{o}, d, f]).$$

7.4.3 *CP-logic*

To represent and reason with probabilistic knowledge, we will use the probabilistic logic language CP-logic [103]. This language is based on the theory of logic programming extended with a probabilistic semantics. The main intuition is that probabilistic logic formulae represent causal rules, that is a logic clause gives a relation from some cause to a set of possible outcomes (each with some probability). A CP-logic program thus describes a causal process.

Definition 7.3. *A* causal probabilistic rule *has the form:*

$$(H_1 : \alpha_1) \; ; \ldots ; (H_n : \alpha_n) \leftarrow B$$

where α_i is the (non-zero) probability of outcome H_i such that $\sum_{i=0}^{n} \alpha_i \leq 1$; H_i are logical atoms and B is the body of the clause.

In other words, a causal rule gives a probability distribution over possible effects of some cause B. Deterministic effects can be modelled simply by a non-probabilistic logic formula $H \leftarrow B$ which is equivalent to $H : 1 \leftarrow B$. Unconditional probabilistic rules can be seen as a prior distribution over logical facts. CP-logic is restricted to finite domains, so although one can write quantified formulae, these are interpreted as a finite set of ground instances.

The probabilistic semantics are based on the work by Shafer [86] on probability trees. The main idea behind this is that probabilistic processes are best described by a dynamic unfolding of events. Each node in the tree represents some state of the domain, transitions between nodes are probabilistic events as described by the causal probabilistic rules in the knowledge base, and each outgoing edge is some alternative outcome labelled with its probability. The leaves of a probability tree each describe a possible outcome of the events modelled in the knowledge base. All events are considered to be independent and dependencies have to be modelled explicitly in the rules. As each causal rule fires independently, it follows that the probability of a leaf node l is the product of the labels on the edges from l to the root of the tree. Since there may be multiple series of events that lead to the same final state the probability of an interpretation is the sum over all the leaves in the tree that share the same interpretation. Each rule is independent, which means that multiple rules with the same outcome are independent causes. In CP-logic, these independent causes are interpreted as a noisy-OR. See Vennekens et al. [103] for details and [50] for alternative interpretations.

Example 7.5. *In Figure 7.2 a CP-logic event tree is shown, representing the situation of whether a COPD patient suffers an exacerbation caused by either an infection or by breathing in a noxious substance. The tree follows from these CP-rules:*

$$\mathsf{exacerbation} : 0.6 \leftarrow \mathsf{infection}.$$

$$\mathsf{exacerbation} : 0.2 \leftarrow \mathsf{noxious\_substance}.$$

$$\mathsf{infection} : 0.05.$$

$$\mathsf{noxious\_substance} : 0.01.$$

The probability of an exacerbation can be computed by summing over the leaves l that contain E (short for exacerbation) on the path from the root to l. Note that on some branches E occurs twice, which is allowed because there are two rules that have E as a consequence. The probability of, for example, the leftmost path is $0.05 \cdot 0.6 \cdot 0.01 \cdot 0.2 = 0.00006$ and the probability of an exacerbation is:

$$0.00006 + 0.00024 + 0.0297 + 0.00004 + 0.0019 = 0.03194.$$

7.4.4 *Markov processes*

Markov processes are commonly used as representation of uncertainty and time. The following definitions will be useful later on. Suppose we have a Markov process represented by the chain

$$X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_t \rightarrow \cdots$$

The joint probability up to some point T is

$$\begin{aligned} P(X_{1:T}) &= P(X_T \mid X_{T-1}, X_{T-2}, \dots) P(X_{T-1}, X_{T-2}, \dots) \\ &= P(X_1) \prod_{t=2}^{T} P(X_t \mid X_{1:t}), \end{aligned}$$

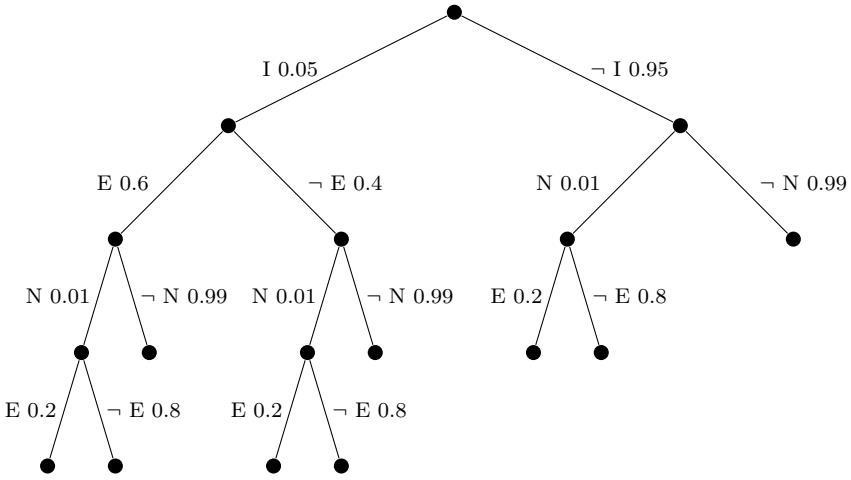


Figure 7.2: A probability tree, where I is short for the *infection* event, N denotes *noxious substance* and E is *exacerbation*.

where the notation $1:t$ is an abbreviation of the sequence from 1 to t. The factorised notation is useful as we are primarily interested in Bayesian network representations. A common independence assumption is the first order Markov property $P(X_{t+1} \mid X_{1:t}) = P(X_{t+1} \mid X_t)$, which states that the future does not depend on the past given the present. The joint probability then simplifies to

$$P(X_{1:T}) = P(X_1) \prod_{t=2}^{T} P(X_t \mid X_{t-1}).$$

A generalisation of Markov chains are dynamic Bayesian networks (DBNs) that allow modelling independences in the state space description. The state space is then described by a graph $G = \langle V, A \rangle$ in which vertices V represent variables X at each time point, arcs A represent dependences either within or between time slices and the joint probability of the process variables X factorises over the graph. The joint probability is given by:

$$\begin{aligned} P(X_{1:T,1:n}) &= P(X_{1,1:n}) \prod_{t=2}^{T} P(X_{t,1:n} \mid X_{t-1,1:n}) \\ &= \prod_{t=1}^{T} \prod_{i=1}^{n} P(X_{t,i} \mid \mathrm{pa}(X_{t,i})) \end{aligned}$$

Note that $\mathrm{pa}(X)$ denotes the parents in the graph of X, which can either be in the same time slice or in time slice $t-1$, again assuming the first order Markov property.

It turns out that we can describe Markov processes in CP-logic as is shown by the work of Thon et al. [93] on CPT-L, which represents observable homogeneous Markov processes in CP-logic.

7.5 A PROBABILISTIC LOGIC OF QUALITATIVE TIME

When modelling real world situations, qualitative time is useful for those processes for which precise timing information is unavailable. However, it may be possible to obtain likelihood information, telling us that some event is more likely to happen at a particular time, even when the timing information is imprecise. This leads to temporal process descriptions – as represented with Allen's logic – extended with probabilistic information. The expressive power of CP-logic will appear sufficient to act as a basis for such an extended, qualitative temporal and uncertain logic.

7.5.1 *A framework for uncertain temporal reasoning*

To model uncertain processes we are primarily interested in the occurrence of events. In our context we consider events that are uniquely associated with intervals, and this association is expressed by means of a time-interval index. We interpret events as taking place throughout their associated intervals.

Now to define temporal events, we index facts with time intervals from the set of all time intervals $\mathcal{I}$.

Definition 7.4. *Let $\mathcal{E}$ denote the* event space *containing all probabilistic events of interest. A* temporal uncertain event *E_I is defined as a probabilistic event $E \subseteq \mathcal{E}$ that occurs in time interval $I \in \mathcal{I}$.*

The Boolean algebra of temporal events $B(\mathcal{E}_\mathcal{I})$, where $\mathcal{E}_\mathcal{I}$ is defined as $\mathcal{E}_\mathcal{I} = \{E_I \mid E \subseteq \mathcal{E}, I \in \mathcal{I}\}$, should obey certain rules, taking into account the time interval indices of the events. The elements of the Boolean algebra are obtained by constructing conjunctions of events $(E_I \wedge E'_J)$, disjunctions $(E_I \vee E'_J)$ and negations $\neg E_I$, with events $E_I, E'_J \in \mathcal{E}_\mathcal{I}$.

With all the basic ingredients defined above, it is now possible to define a framework for uncertain temporal reasoning.

Definition 7.5. *A* probabilistic qualitative time algebra *$\mathcal{PQT}$ is defined as a 4-tuple $\mathcal{PQT} = \langle \mathcal{I}, \mathcal{E}_\mathcal{I}, \mathcal{B}, P \rangle$, where*

- *$\mathcal{I}$ is the set of all time intervals;*
- *$\mathcal{E}_\mathcal{I}$ is the set of temporal uncertain events;*
- *$\mathcal{B}$ is the set of basic Allen relations;*
- *P is an associated joint probability distribution.*

Below, we will use CP-logic as a practical language to implement the framework.

7.5.2 *On events and intervals*

There are certain properties of temporal events which require further attention, and which will be considered subsequently.

For $E = E'$ with temporal events E_I and E'_J, there is an interaction between the Boolean operations on temporal events and the Allen relationships of the time intervals. For example, when IbJ holds, then $(E_I \wedge E_J)$ cannot be simplified; however, if $IeqJ$ holds, then $(E_I \wedge E_J) = E_I = E_J$. In addition, when two intervals I and J meet, i.e. ImJ holds, then $(E_I \wedge E_J) = E_K$ with $K = I \cup J$. This expresses that event E actually occurred during interval K.

Proposition 7.1. *For intervals $I, J \in \mathcal{I}$ and a relation IRJ it holds that*

$$(E_I \wedge E_J) = E_{I \cup J} \tag{7.1}$$

if $R \in \mathcal{B} \setminus \{b, \bar{b}\}$.

In other words, it holds for all cases except when I and J are disconnected intervals.

Proof. The proposition can be understood by splitting the intervals into the intersection $K = I \cap J$, and the remainders $I' = I \setminus K$ and $J' = J \setminus K$. Note that $\{b, \bar{b}, m, \bar{m}\}$ are the only relations for which the intersection K is empty, but that for $\{m, \bar{m}\}$ we have that $\sup I = \inf J$ and hence I, J are connected. Then by realising that $(I' \cup K) \cup (K \cup J') = I' \cup K \cup J'$ and that the same event in connected intervals cannot be distinguished, we obtain $E_{I \cup J}$. □

Note that Equation (7.1) is true for any temporal relation R if we define the time-interval index I of a temporal event E_I as a set of time intervals, rather than as a single time interval, with the singleton set as a special case. This would generalise Allen's algebra of temporal relations to relations between sets of time intervals. As it is not our intention in this paper to change the algebraic basis of Allen's algebra, we will not pursue this idea further.

For a disjunction $E_I \vee E_J$, we can apply the results from the proof of Proposition 7.1 to obtain that

$$(E_I \vee E_J) = E_K \wedge (E_{I'} \vee E_{J'}) \tag{7.2}$$

where K, I' and J' are defined as in the proof above, for IRJ, with $R \in \mathcal{B} \setminus \{b, \bar{b}\}$.

7.5.3 *Uncertainty*

There are various ways in which uncertainty can be incorporated into the language:

- Uncertainty can be expressed with respect to the temporal events E_I.
- Uncertainty can be expressed with regard to the relation between time intervals IRJ.

When combined for two temporal events E_I and E'_J, these assumptions give rise to a joint probability distribution of the form

$$P(E_I, E'_J, IRJ),$$

with R an Allen relation. Note that this could also be the special case where we have a single event E associated with both interval I and interval J.

Often one is interested in more than two events, which can be modelled with a joint probability over all the events and pairwise temporal relations between the events. For example for three events we have the joint probability

$$P(E_I, E'_J, E''_K, IRJ, IR'K, JR''K).$$

Sometimes this joint probability will simplify due to independences between the events. In the following we focus on situations with two events, yet similar arguments can also be made for multiple events.

Certain Allen relations

By conditioning on IRJ, one removes part of the uncertainty, yielding:

$$P(E_I, E'_J \mid IRJ).$$

In this case, only events are uncertain, and relations between interval are considered to be part of the (deterministic) logic specification. We can now define a distribution over temporal events.

Definition 7.6. *The probability of a Boolean expression of events is given by the probability function $P : B(\mathcal{E}_\mathcal{I}) \to [0, 1]$, where $B(\mathcal{E}_\mathcal{I})$ denotes the Boolean algebra over the set of temporal events $\mathcal{E}_\mathcal{I}$.*

Here uncertainty is introduced at the level of events attached to particular intervals, which means we are looking at what happens at a greater level of detail than the qualitative relations describe. However, the nature of Allen's qualitative algebra dictates that relationships between the intervals are more important than the exact intervals. Consequently, the following parameter invariance appears appropriate to restrict arbitrary models to those models where the qualitative relations are the key primitives for temporal uncertainty:

$$\forall R, I, J, K, L : P(E_I, E'_J \mid IRJ) = P(E_K, E'_L \mid KRL), \tag{7.3}$$

i.e. when the Allen relation R between (potentially) different intervals is the same, the probability of joint occurrence of the associated temporal events is also the same. This invariance ensures that the influence of time on probabilities of events is governed by the temporal relations.

This invariance is actually related to a common assumption for dynamic Bayesian networks, where one assumes parameter invariance over time to limit the number of parameters. There is however a subtle difference. The invariance that is commonly assumed for DBNs is that $P(X_t \mid X_{t-1})$ is equal for all t. The invariance in Equation (7.3) however states that as long as two events are related through the same temporal relation their probabilities are equal, which is not about repetition invariance. Yet the relation with Markov processes is interesting and we will now show that the qualitative time framework can be used to model Markov processes.

The important properties of a Markov process as described in Section 7.4.4 for a representation in terms of Allen's relations are the ordering of events and the factorisation over time slices. Hence a chain of events, connected with the relations $R_i \in \{b, m\}$ represents the correct ordering:

$$E_I \ R_1 \ E'_J \ R_2 \ E''_K \ R_3 \ \cdots$$

Since this results in disjoint time slices, the factorisation property also holds under the first order Markov assumption.

To generalise to representing a DBN with Allen relations we have to somehow map events to time slices depending on the relations between them. A division in classes of relations turns out to be useful here:

$$\mathcal{B} = \mathcal{P} \cup \bar{\mathcal{P}} \cup \mathcal{C} \cup \mathcal{O}$$

where the subsets are the precedence relations $\mathcal{P} = \{b, m\}$, the inverse precedence relations $\bar{\mathcal{P}} = \{\bar{b}, \bar{m}\}$, the concurrent relations $\mathcal{C} = \{s, \bar{s}, d, \bar{d}, f, \bar{f}, eq\}$ and the overlap relations $\mathcal{O} = \{o, \bar{o}\}$.

Definition 7.7. *A* temporal partition Π *of a set of temporal events $\mathcal{E}_\mathcal{I}$ is a partition such that for all sets $L \in \Pi$ it holds that: $E_I \ R \ E'_J$ with $R \in \mathcal{C}$ iff $E_I, E'_J \in L$; and $E_I \ R \ E'_J$ with $R \in \mathcal{P} \cup \bar{\mathcal{P}}$ iff $E_I \in L$, $E'_J \in M$, $M \in \Pi$ and $L \neq M$.*

First, we need to show that the uncertain temporal events can be linearly ordered. Recall that a linear order $\leq$ is transitive, antisymmetric and total.

Lemma 7.1. *Let $\mathcal{E}_\mathcal{I}$ be a set of temporal events and Π a temporal partition of $\mathcal{E}_\mathcal{I}$, then, the temporal events can be linearly ordered.*

Proof. The sets $L \in \Pi$ are linearly ordered by the fact that the precedence relation induces a linear order $\leq$. Hence, temporal events in different sets are also linearly ordered. Events in each set L are equivalent as they are neither $<$ nor $>$, showing antisymmetry. Between any two temporal events in a temporal partition a relation in $\mathcal{B} \setminus \mathcal{O}$ holds, so totality follows. □

From the linear order that can be imposed on the temporal events it follows that one can map temporal events to a DBN.

Proposition 7.2. *A set of temporal events $\mathcal{E}_\mathcal{I}$ with a temporal partition Π can be mapped to a DBN with the first order Markov assumption.*

Proof. A DBN with the first order Markov assumption has arcs either within a time slice or from time slice t to $t+1$. The temporal partition results in a linear order by Lemma 7.1. Now let each $L \in \Pi$ correspond to a time slice. Two time slices are connected by arcs in the DBN graph if $L < M$ with $L, M \in \Pi$, $E'_J \; R \; E_I$ holds with $R \in \bar{\mathcal{P}}$, $E'_J \in M$ and $E_I \in L$, and there is no set $N \in \Pi$ such that $L < N < M$. □

In other words, we can construct a first order Markov DBN for those temporal networks that only contain relations in $\mathcal{C}$ and $\mathcal{P}$, resulting in intra and inter time-slice arcs respectively. If we also want to take into account the overlap relations $I \; \mathcal{O} \; J$ the construction is less straightforward. The overlapping events can be considered an indivisible unit, which allows us to assign the combined interval $I \cup J$ to a single time slice. An undesirable consequence is that a chain of overlapping events would result in a single time slice for the complete chain. However we can give a restriction that allows us to deal with some overlapping events.

Proposition 7.3. *Let $E_I \; \mathcal{O} \; E'_J$, $E_I, E'_J \in \mathcal{E}_\mathcal{I}$. $\mathcal{E}_\mathcal{I}$ corresponds to a DBN if for all $E''_K \in \mathcal{E}_\mathcal{I}$ we have $E''_K \; R \; E_I$ and $E''_K \; R' \; E'_J$ with $R, R' \in \mathcal{C}$; or R, R' either both in $\mathcal{P}$ or both in $\bar{\mathcal{P}}$.*

Proof. If $R, R' \in \mathcal{C}$ the joined interval $I \cup J$ is also only related via $\mathcal{C}$, which by Proposition 7.2 results in a single time slice. If $R, R' \in \mathcal{P}$ or $R, R' \in \bar{\mathcal{P}}$ the joined interval will by the same proposition result in a separate time slice for the events. □

Uncertain Allen relations

The joint probability $P(E_I, E_J, IRJ)$ represents the combined uncertainty in events and their temporal relations. Quantifying uncertainty in the relations by means of probability generalises the use of disjunctions in Allen's algebra to model for example uncertainty in the ordering of events. Conditioning the joint probability on the temporal relation, we obtain:

$$P(E_I, E_J, IRJ) = P(E_I, E_J \mid IRJ)P(IRJ).$$

For fixed intervals I, J, the following holds:

$$P(I\mathcal{B}J) = P(\bigvee_{r \in \mathcal{B}} IrJ) = \sum_{r \in \mathcal{B}} P(IrJ) = 1,$$

which is due to the mutual exclusivity of the basic relations. Thus, for any subset $R \subseteq \mathcal{B}$ it holds that:

$$P(IRJ) = P(\bigvee_{r \in R} IrJ) = \sum_{r \in R} P(IrJ).$$

So although the set $\mathcal{A}$ is large ($|\mathcal{A}| = 2^{13}$) the probability of a relation $R \in \mathcal{A}$ is simply the sum over the basic relations in R.

Closure operations

When using probabilistic relations it is natural to think about whether the properties of Allen's algebra can be generalised. Returning for a moment to relations instead of a logic representation, we should reconsider the operations from Definition 7.2: *inverse*, *intersection* and *composition*. In [82] some methods were proposed, explained below. However, in our opinion the motivation is lacking and we propose an alternative for the probabilistic equivalent of the intersection operation.

We first introduce the notation $P[I\mathcal{B}J]$ to define a distribution over the elements of $\mathcal{B}$. So for example if for IRJ we have $P(IeqJ) = 0.3$, $P(IbJ) = 0.7$ and $P(IrJ) = 0$ for all $r \in \mathcal{B} \setminus \{eq, b\}$; we write

$$P[I\mathcal{B}J] = [eq : 0.3, b : 0.7].$$

Now, let the following probability distribution for $JR'K$ be given:

$$P[J\mathcal{B}K] = [d : 0.5, o : 0.5].$$

We can now define the probabilistic operations.

INVERSE The inverse of $P[I\mathcal{B}J]$ is found by elementwise inversion $P[J\mathcal{B}I] = [eq : 0.3, \bar{b} : 0.7]$.

COMPOSITION If we compute the composition $I(R \circ R')K$ we obtain $I\{bmosd\}K$, for which we can compute the probabilities via the procedure in [82]

$$\begin{aligned} P(d) &= eq \circ d = 0.3 \cdot 0.5 = 0.15 \\ P(o) &= eq \circ o = 0.3 \cdot 0.5 = 0.15 \\ P(b) &= \ \ b \circ d = 0.7 \cdot 0.5 = 0.35 \\ P(bmosd) &= \ \ b \circ o = 0.7 \cdot 0.5 = 0.35 \end{aligned}$$

Note that $P(Ir''K) = P(IrJ)P(Jr'K)$ with $r'' = r \circ r'$. The composition $b \circ o$ results in multiple relations, which means that the probability 0.35 has to be divided over $\{bmosd\}$. A non-informative choice would be a uniform assignment, but in general we have

$$\begin{aligned} P(I\{b, o, d, (b \vee m \vee o \vee s \vee d)\}K) &= 1 \\ 0.15 \leq P(IdK) &\leq 0.15 + a \\ 0.15 \leq P(IoK) &\leq 0.15 + b \\ 0.35 \leq P(IbK) &\leq 0.35 + c \\ P(ImK) + P(IsK) + a + b + c &= 0.35 \end{aligned}$$

which implies $a, c, b \in [0, 0.35]$.

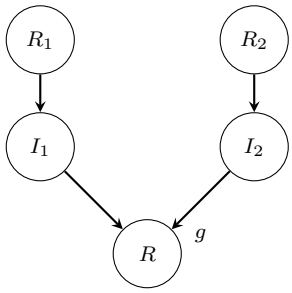


Figure 7.3: Causal independence model to combine the distributions over the relations R_1,R_2 via the interaction function g.

INTERSECTION For the intersection operation we first consider conditioning. Assume that we have the following probability distribution:

$$P[I\mathcal{B}K] = [m : 0.3, o : 0.7]$$

When we now also learn that $I\{m, s\}K$ holds this means that according to Allen's algebra only ImK can hold as $\{m\} = \{m, o\} \cap \{m, s\}$. Probabilistically, this can be viewed as a conditioning event which leads to the conclusion that *meets* holds almost surely, as follows:

$$P(ImK \mid I\{m, s\}K) = \frac{P(ImK)}{P(I\{m, s\}K)} = \frac{P(ImK)}{P(ImK) + P(IsK)} = 1$$

as $P(IsK) = 0$, and

$$P(IoK \mid I\{m, s\}K) = \frac{P(\bot)}{P(I\{m, s\}K)} = 0$$

Here we assume that we know the initial distribution and obtain additional information on which we can condition. In this particular situation we can solve the problem and compute the probabilities as in the example. This is different from the approach taken in [82], where two distributions $P_{R'}(IR'J)$ and $P_{R''}(IR''J)$ are combined into a single distribution $P(IRJ)$:

$$\forall x \in R : P_R(IxJ) = \frac{1}{Z} \frac{P_{R'}(IxJ)P_{R''}(IxJ)}{P_{R'}(IxJ) + P_{R''}(IxJ)},$$

where Z is the normalisation $Z = \sum_{z \in R} P_R(IzJ)$. There appears to be little theoretical justification for this approach however. Instead, we opt to use the framework of causal independence, which offers a principled approach to model interactions between probabilistic events. Modelling the intersection operation then results in the graphical model shown in Figure 7.3.

The behaviour of the causal independence model depends on the choice of interaction function, g in the figure. Using some variant of the noisy-AND appears appropriate, due to the similarity to the logical case. Since the relations are not binary valued the standard approach does not apply, but the generalisation by Srinivas [90] offers an alternative. The probability of R given R_1 and R_2 is given by:

$$P(R \mid R_1, R_2) = \sum_{\{I_1, I_2 \mid g(I_1, I_2) = R\}} \prod_i P(I_i \mid R_i)$$

where the probabilities $P(I_i \mid R_i)$ are given by:

$$P(I_i = x \mid R_i = y) = \begin{cases} P(\text{pass}(I_i)) + P(\text{fail}(I_i = x)) & \text{if } x = y \\ P(\text{fail}(I_i = x)) & \text{otherwise} \end{cases}$$

where $\text{fail}(I_i = x)$ indicates that I_i takes on the value x irrespective of the value of R_i, whereas $\text{pass}(I_i)$ indicates that no failure occurs and hence

$$P(\text{pass}(I_i)) = 1 - \sum_{z \in \text{val}(R_i)} P(\text{fail}(I_i = z)).$$

For the interaction function g an analogue to the conjunction used in Allen's algebra is the componentwise AND-function

$$R = g(I_1, I_2) \text{ with } R(i) = I_1(i) \wedge I_2(i),$$

where the components correspond to the basic relations $\mathcal{B}$ and $R(i)$ denotes the ith component of R and similarly for I. This then results in the probabilities

$$P(R = r) = P(g(I_1, I_2) = r) = P(I_1 = r)P(I_2 = r),$$

which is not a normalised distribution, so after normalisation we obtain

$$P(R = r) = \frac{1}{Z} P(g(I_1, I_2) = r) = \frac{1}{Z} P(I_1 = r)P(I_2 = r),$$

where $Z = \sum_z P(I_1 = z)P(I_2 = z)$.

7.5.4 *The framework in CP-logic*

Probabilistic logic, and specifically CP-logic, offers a way to implement the framework described above. In CP-logic events are represented by facts, which are interpreted as independent events. Relations between facts are stated by logical expressions and through these expressions, dependences between events embedded in the facts can be introduced.

We can now define our language which we call *Qualitative Time CP-logic* as a restriction of CP-logic. The language consists of CP-rules (Definition 7.3), where the predicates model temporal probabilistic events (Definition 7.4). Furthermore, the special predicates $\mathsf{r}/3$ and $\mathsf{allen}/3$ (defined in Section 7.4.2), are used to denote the temporal relations between the events in each rule. We allow abbreviations where the omission of an interval means that the event is static over the time of interest and the omission of a relation in a rule indicates that $\mathsf{allen}(\cdot, \cdot, [\mathcal{B}])$ holds. These abbreviations allow us to incorporate specific static rules in standard CP-logic syntax. The probabilistic semantics are inherited from CP-logic, which leads to the result in the following subsection.

7.5.5 *Reasoning with temporal events*

Now that we have defined uncertainty with respect to temporal events we can examine reasoning with uncertainty and time. To capture the qualitative nature of the time representation in logic, we represent intervals as atoms, foregoing the need to reason about the underlying time points.

The probability of temporally related events is of interest, which is, as we have seen, the probability $P(E_I, E'_J, IRJ)$. The following proposition shows how we can compute this probability via the CP-logic semantics, given that IRJ is a logical relation between the intervals associated with the events.

Proposition 7.4. *Let R be a relation $R \in \mathcal{A}$, and $E_I, E'_J \in \mathcal{E}_\mathcal{I}$ events; the probability $P(E_I, E'_J, IRJ)$ is then obtained as follows:*

$$\begin{aligned} P(E_I, E'_J, IRJ) &= P(E_I \mid E'_J, IRJ) P(E'_J \mid IRJ) P(IRJ) \\ &= \sum_\ell \mathbb{I}_\ell(E_I, E'_J, IRJ) \prod_{e_\ell} P(e_\ell) \end{aligned}$$

where ℓ denotes a leaf node in the tree, e_ℓ is an edge on the path from ℓ to the root of the tree; and $\mathbb{I}_\ell(x)$ is an indicator function that is 1 if x is true in ℓ.

Proof. (sketch) The proposition follows from the probability tree semantics of CP-logic. An event E_I is either true with a priori probability p, or is the consequent of a rule $E_I : p \leftarrow B$, with B the body of the clause. A rule decomposes the joint probability $P(H, B)$ into $P(H \mid B)P(B)$. The probability of a particular event can hence be computed via a series of rule applications, which can be represented as a event tree. As CP-rules fire independently, we obtain the product from the proposition as a path in the tree. As there are multiple series of rules (logic proofs) that can lead to the same events, the total probability $P(E_I, E'_J, IRJ)$ is the sum over those leaves ℓ in which the events are true. □

The proof relies on properties of CP-logic, the details of which can be found in [103]. Note that the temporal relations can be seen as a special kind of event. We use them only as constraints in the body of rules, but this is not a technical requirement.

Example 7.6. *We are again interested in modelling the relation between an infection and the occurrence of an exacerbation. Given the temporal relations* overlaps *and* equals *we now obtain the CP-logic rules:*

$$\mathsf{exacerbation}(j) : 0.7;\ \neg\mathsf{exacerbation}(j) : 0.3 \leftarrow \mathsf{infection}(i), \mathsf{allen}(i, j, [o, eq]).$$
$$\mathsf{r}(i, j, o) : 0.8;\ \mathsf{r}(i, j, eq) : 0.2.$$
$$\mathsf{infection}(i) : 0.01.$$

It follows that the probability of an exacerbation is a priori $7 \cdot 10^{-3}$, and if we would have (definite) evidence for an infection the probability would be 0.7*.*

The distribution over possible temporal relations in this case has the same result as when $\mathsf{r}(i, j, o)$ or $\mathsf{r}(i, j, eq)$ holds with probability 1, since the same rule for exacerbation fires in both cases. It does show however that by specifying a distribution in this way the relations are mutually exclusive also in the reasoning process, which ensures that the right probabilities are computed, as given in Proposition 7.4.

7.5.6 *Metric relations*

A number of extensions to Allen's algebra have been proposed (e.g. [29, 53]) based on various kinds of constraint reasoning. The added value lies mostly in metric constraints, which allow numerical constraints in addition to the pure qualitative relations in Allen's algebra. Jonsson and Bäckström [52] proposed the framework of disjunctive linear relations, which supports tractable reasoning with metric relations, while being at least as expressive as a number of earlier proposals (see also [58] on tractability results).

For practical medical applications metric constraints seem a useful addition to qualitative relations, enabling a more precise model of the temporal relation in those cases where the necessary level of quantitative detail is attainable. The essence of adding metric information is that additional constraints between interval endpoints should

be satisfied, which can be modelled in the logic part of our specification. That is, special instances of temporal relations can be defined containing a metric constraint. The metric extension is useful to model relations that cannot be expressed within the original algebra, for instance a physical constraint that some effect occurs only after a certain amount of time.

Formally, for a qualitative temporal relation $R \in \mathcal{B}$, a quantified version $R_m(x, y)$ of this relation has numerical parameters x, y that bound the distance between interval endpoints. Depending on the temporal relation, this definition can still be implemented in different ways. For instance a constraint like 'I is at least two time points before J' can be specified with a single quantitative parameter, leading to an alternative definition of *before*:

$$b_m(I, J, X, Y) \leftarrow I^+ + X < J^-,$$

meaning that I is at least X time points before J. Here we did not need the parameter Y, in practice different types of metric constraints might be useful that do require two parameters, like 'I between 2 and 4 time points before J':

$$b_m(I, J, X, Y) \leftarrow X \leq J^- - I^+ \leq Y.$$

With this parametrisation we can construct a quantified set of temporal relations on a level of detail that is appropriate for the domain.

To specify a distribution over relations, notice that the metric versions of the qualitative relations define a subset of the intervals, i.e. in terms of relations as sets $b_m \subseteq b$, hence the probability assigned to before can be distributed over the metric relations under the condition that they are disjoint. For example, if in the original distribution the probability of *before* was p and we want to introduce the relation 'at least two time points before', $b_{>2}$, with probability $\frac{1}{2}p$, it suffices to define $b_{>2}$ and add the rule

$$Ib_{>2}J : 0.5 \leftarrow IbJ.$$

7.6 VALIDATION OF THE QUALITATIVE PROBABILISTIC FRAMEWORK

In this section, we return to the two typical biomedical cases described in Section 7.2 and we use them here to validate the framework. We investigate whether the framework has sufficient expressive power to capture the biomedical knowledge concerning those cases. As the same framework is used to handle different cases, this shows that the framework is generic and has potential for dealing with a wide variety of other biomedical problems as well.

7.6.1 *Temporal Nodes Bayesian Networks*

We first study a particular class of temporal models called temporal nodes Bayesian networks (TNBN) [6], which represent event intervals as values of variables in a Bayesian network. This allows reasoning about events in those situations where one is not interested in dynamic changes over time but in single occurrences that are temporally uncertain. This class of models describes a worthwhile subset of the situations that we want to be able to represent.

In a recent paper [48] a TNBN was used to model interactions between antiretroviral drugs and mutations of the Human Immunodeficiency Virus, HIV, that result in drug resistance. The temporally uncertain aspect is here the occurrence of the mutations given other mutations and an antiretroviral drug or a cocktail of those drugs (which present an evolutionary pressure for particular mutations to succeed).

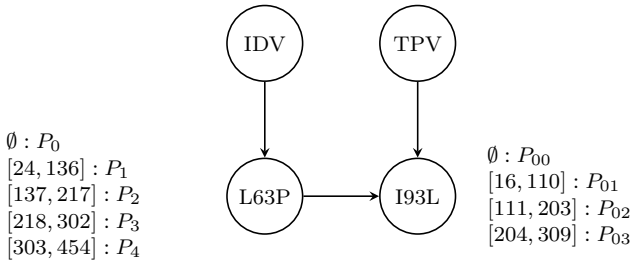


Figure 7.4: Submodel of the TNBN from [48].

The presence of drugs is modelled atemporally, while each mutation has a probability of occurrence in each of a set of consecutive intervals.

To represent this in our logic we first need to represent the atemporal information, which consists of a set of ground clauses of the type:

$$\mathsf{anti\_viral}(X) : P.$$

indicating that a particular drug X occurs with probability P, although when studying a particular situation it will typically be known which drugs are present, allowing us to instantiate these clauses to either true of false. Each mutation is influenced by other mutations and drugs resulting in rules of the type:

$$\mathsf{mutation}(X, \emptyset) : P_0;\ \mathsf{mutation}(X, i_1) : P_1;\ \mathsf{mutation}(X, i_2) : P_2;$$
$$\ldots; \mathsf{mutation}(X, i_n) : P_n \leftarrow \mathsf{mutation}(Y, j), \ldots, \mathsf{anti\_viral}(Z), \ldots$$

where $i_1, i_2, \ldots, i_n, j$ are ground intervals, $\emptyset$ indicates the event did not occur and $\sum_{k=0}^{n} P_k = 1$ holds as usual. In terms of Allen's relations we have the constraint

$$i_1\ m\ i_2\ m \cdots m\ i_n,$$

for all temporal variables. However, because each variable is allowed to take on different interval values, interactions between mutations result in other temporal relations. In TNBNs these relations are modelled implicitly in the probability distributions. That is, if $\mathsf{mutation}(X, I) \leftarrow \mathsf{mutation}(Y, J)$, the $n \cdot (m+1)$ parameters for i_k, j_l with $k \in \{0, \ldots, n\}, l \in \{0, \ldots, m\}$ encode the Cartesian product of possible combinations of I and J.

Concretely, we model the two most common mutations, L63P and I93L, and two drugs, IDV and TPV, that influence them. This is a submodel of one of the networks described in [48], reproduced in Figure 7.4. This results in the following rules:

$$\mathsf{anti\_viral}(\mathrm{IDV}).$$
$$\mathsf{anti\_viral}(\mathrm{TPV}).$$
$$\mathsf{mutation}(\mathrm{L63P}, \emptyset) : P_0;\ \mathsf{mutation}(\mathrm{L63P}, [24, 136]) : P_1;$$
$$\mathsf{mutation}(\mathrm{L63P}, [137, 217]) : P_2;\ \mathsf{mutation}(\mathrm{L63P}, [218, 302]) : P_3;$$
$$\mathsf{mutation}(\mathrm{L63P}, [303, 454]) : P_4 \leftarrow \mathsf{anti\_viral}(\mathrm{IDV}).$$
$$\mathsf{mutation}(\mathrm{I93L}, \emptyset) : P_{00};\ \mathsf{mutation}(\mathrm{I93L}, [16, 110]) : P_{01};$$
$$\mathsf{mutation}(\mathrm{I93L}, [111, 203]) : P_{02};\ \mathsf{mutation}(\mathrm{I93L}, [204, 309]) : P_{03};$$
$$\leftarrow \mathsf{anti\_viral}(\mathrm{TPV}),\ \mathsf{mutation}(\mathrm{L63P}, \emptyset).$$

where for brevity we have omitted the rules for the other parent configurations of I93L which are analogous to the one shown. It is easy to see that the probability of for example $\mathsf{mutation}(\text{I93L}, \emptyset)$ can be computed as:

$$\begin{aligned} &P(\mathsf{mutation}(\text{I93L}, \emptyset)) \\ &\quad = \sum_{I} P(\mathsf{mutation}(\text{L63P}, I)) P(\mathsf{mutation}(\text{I93L}, \emptyset) \mid \mathsf{mutation}(\text{L63P}, I)) \\ &\quad = \sum_{i=0}^{4} P_i P_{i0}. \end{aligned}$$

This is the expected result of the CP-logic semantics and of TNBNs, and indeed our rule representation is equivalent to the TNBN.

With our representation we can also quite simply introduce extensions with little additional work, as the framework takes care of the probabilistic reasoning. For example it may be interesting to make the temporal representation more qualitative, instead of the concrete intervals used now. To do so we can model the interaction between mutations by means of a temporal relation, and especially *before* is a natural candidate. For instance, the influence of L63P on I93L can be modelled as two rules, for the case that L63P does not occur (not shown) and for the case that L63P occurs before I93L:

$$\begin{aligned} &\mathsf{mutation}(\text{I93L}, \emptyset) : P_{10};\ \mathsf{mutation}(\text{I93L}, J) : P_{11}; \\ &\qquad \leftarrow \mathsf{anti\_viral}(\text{TPV}),\ \mathsf{mutation}(\text{L63P}, I),\ \mathsf{r}(I, J, b). \end{aligned}$$

Another possible extension would be to also take into account time for the antiviral drugs. It appears plausible that the influence of a drug is concentrated in the interval in which it is present, with possibly limited influence some time longer due to the fact that it changed the population of viruses. We can capture this idea in the rules:

$$\begin{aligned} &\mathsf{anti\_viral}(\text{IDV}, I). \\ &\mathsf{mutation}(\text{L63P}, \emptyset) : P_0;\ \mathsf{mutation}(\text{L63P}, J) : P_1; \\ &\qquad \leftarrow \mathsf{anti\_viral}(\text{IDV}, I),\ \mathsf{r}(I, J, eq). \\ &\mathsf{mutation}(\text{L63P}, \emptyset) : P_0';\ \mathsf{mutation}(\text{L63P}, J) : P_1'; \\ &\qquad \leftarrow \mathsf{anti\_viral}(\text{IDV}, I),\ \mathsf{r}(I, J, m_x). \end{aligned}$$

where the relation m_x is a metric version of the *meets* relation, defined as $I m_x J \leftarrow I^+ = J^- \wedge J^+ - J^- \leq x$, allowing us to limit the time of influence of the drug after it is no longer administered.

The ease with which we can incorporate ideas for extensions in the language shows the versatility of our framework. The logic offers a readable language to represent relations both atemporal and temporal which together with the probabilistic semantics results in a compelling framework to represent many clinical situations. We now return to the problem of modelling the disease process of COPD.

7.6.2 *Describing the evolution of COPD*

Our framework is particularly well-suited to the clinical task of monitoring patients with chronic obstructive pulmonary disease. As the previous examples have shown both the time and uncertainty aspects play a role in COPD management. In this section we develop a model of COPD disease progression in more detail.

We start with a clinical description of influence between the relevant variables and their qualitative temporal relations. Relevant events to construct a process

description of COPD progression are the outcome variable *exacerbation* its usual cause *infection* and the observable symptoms, signs and lab data. In the context of monitoring patients at home, access to signs or lab data is limited and we focus on symptoms, specifically dyspnea (difficulty in breathing), sputum colour and volume, cough and reduced activity. The patients from whom the data have been acquired typically have around two exacerbations per year, and consequently exacerbation symptoms will be largely absent most of the time. The situations we are interested in are the episodes in which an exacerbation occurs. When an exacerbation develops, different symptoms will increase over time until either through medication or through natural recovery the symptoms return to normal levels.

After having identified the variables our logic representation makes it is easy to model explicitly the structural relations in the domain. That is, we construct rules of the type:

$$\mathsf{symptom}(P, S) \leftarrow \mathsf{infection}(P, Y), \mathsf{symptom}(S), \mathsf{patient}(P).$$
$$\mathsf{symptom}(dyspnea).$$
$$\mathsf{symptom}(cough).$$
$$\mathsf{symptom}(sputum).$$
$$\mathsf{symptom}(activity).$$

This indicates that a patient P with an infection Y (which would allows us to specify different kinds of infections, e.g. bacterial or viral; or even more detailed by identifying specific bacteria that often cause airway infections like *H. Influenzae*, or *S. Pneumoniae*, but here we limit ourselves to just modelling the presence of an infection), has symptoms from the set $\{dyspnea, cough, sputum, activity\}$. In addition we should take into account interdependencies of the symptoms, for example, that an increase in dyspnea can cause a decrease of activity (the fact that some symptoms indicate an increase and others a decrease can easily be represented explicitly in the logic but is omitted for simplicity):

$$\mathsf{symptom}(P, activity) \leftarrow \mathsf{symptom}(P, dyspnea), \mathsf{patient}(P).$$

The resulting rules are quite readable, and both obtaining and verifying the domain knowledge with help from an expert is a realistic option.

Given the structural properties the next step is modelling the temporal characteristics. It is unrealistic to have precise timing information as this depends on many patient specific and environmental factors. Allen's algebra is a useful compromise between unattainable precise temporal knowledge and disregarding the temporal dimension of the problem completely. Concretely, we can describe the temporal structure of the process by assigning relations between infection events, exacerbation events and symptoms, for example in a substantial number of cases the first symptom to increase is dyspnea, which often stays present until the exacerbation ends. So the most likely temporal relation would be:

$$\mathsf{infection}\ \{\bar{d}o\}\ \mathsf{dyspnea}\ \{\bar{f}o\}\ \mathsf{exacerbation}.$$

Similarly we can use domain knowledge to define relations for the other symptoms with respect to infection and exacerbation events. Analogous to the structural relations, we can also take into account the temporal relations between symptoms. It should be noted however that if the desired information is hard to obtain, this just means that we have no explicit constraint on the temporal relation between events, which means that all basic relations are possible.

In general it will often not be easy to specify structural and temporal characteristics exactly, requiring the inclusion of uncertain information. The advantage of

starting with logic (without probabilities) is that the main structure of the process can be determined more easily. The choice for a probabilistic logic like CP-logic allows us to subsequently refine the representation within the same framework to include uncertainty. If clinical data is available, it can be applied in this stage to estimate probabilities.

The temporal relations may also be underspecified which can be modelled with a distribution over the basic temporal relations. For example derived from relative frequencies in the data and a knowledge-based prior distribution. We then might obtain for example the following distribution for the relation between an exacerbation and dyspnea – which in CP-logic notation consists of a set of statements like $\mathsf{r}(I, J, X) : P$, where X is a temporal relation and P a probability – shown here more compactly:

$$\begin{array}{lllllll} b\ 0 & m\ 0 & o\ 0 & s\ 0.01 & d\ 0.05 & f\ 0.35 & eq\ 0.18 \\ \bar{b}\ 0 & \bar{m}\ 0 & \bar{o}\ 0.2 & \bar{s}\ 0.05 & \bar{d}\ 0.1 & \bar{f}\ 0.05 & \end{array} \tag{7.4}$$

The relation *before* and *meets* have been assigned probability zero, because we are only interested in symptoms that actually co-occur with an exacerbation. And *overlaps* has zero probability since it makes little sense to consider exacerbations that start before their symptoms.

Putting it all together we end up with a representation that includes structural, temporal and uncertainty information. To give an example of the reasoning in our COPD model we restrict ourselves to the symptoms *dyspnea* and *cough*, their cause *infection* and the effect *exacerbation*. We assume the temporal relation *equals* for the relation between the symptoms and infection. Then using the predicates we introduced in Section 7.4.2 and somewhat more compact notation for readability (omitting the specification of the patient P, writing symptoms as predicates and some set-notation in 7.8), the rules to model this process are the following:

$$\mathsf{dyspnea}(I) : 0.8 \leftarrow \mathsf{infection}(J), \mathsf{r}(I, J, eq). \tag{7.5}$$

$$\mathsf{cough}(I) : 0.6 \leftarrow \mathsf{infection}(J), \mathsf{r}(I, J, \bar{d}). \tag{7.6}$$

$$\mathsf{exacerbation}(I) : 0.95 \leftarrow \mathsf{dyspnea}(J), \mathsf{allen}(I, J, [\bar{o}, f]), \mathsf{cough}(K), \mathsf{r}(I, K, X). \tag{7.7}$$

$$\mathsf{exacerbation}(I) : 0.7 \leftarrow \mathsf{dyspnea}(J), \mathsf{allen}(I, J, [\mathcal{C} \setminus \{f\}]), \mathsf{cough}(K), \mathsf{r}(I, K, X). \tag{7.8}$$

$$\mathsf{infection}(I) : 0.05 \leftarrow \mathsf{infection}(J), \mathsf{r}(I, J, \bar{b}). \tag{7.9}$$

$$\mathsf{infection}(i1) : 0.05. \tag{7.10}$$

Both exacerbations and individual symptoms are recurring events but the start event of a series of occurrences is the only event for which we need to model the repetition. This is expressed by rule 7.9. To predict the probability of an exacerbation in a particular interval, we need probabilities for the dependencies between exacerbations and symptoms given a temporal relation. Instead of exhaustively enumerating all relations, we identify the dominant pattern: exacerbation is *overlapped-by* or *finishes* dyspnea. The only probabilities that have to be specified are those for the dominant temporal pattern (rule 7.7) and for the case summarising the other situations (rule 7.8). Instantiations of the relation predicates for specific intervals are omitted for conciseness.

With this representation we can now query probabilities of interest given evidence (observations). For example, let us assume the relations between infection and the symptoms, $\mathsf{r}(\mathsf{dyspnea}, \mathsf{infection}, eq)$ and $\mathsf{r}(\mathsf{cough}, \mathsf{infection}, \bar{d})$ hold. The distribution for $\mathsf{r}(\mathsf{exacerbation}, \mathsf{cough}, X)$ can be derived from $\mathsf{r}(\mathsf{dyspnea}, \mathsf{infection}, eq)$,

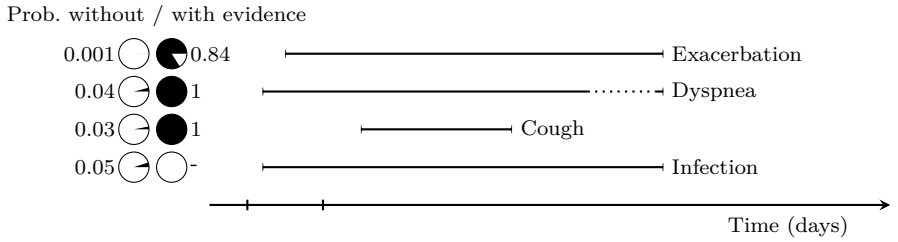


Figure 7.5: Graphical representation of the example time-course and probabilities. The proportion of the circle that is shaded indicates the probability.

$\mathsf{r}(\mathsf{infection}, \mathsf{cough}, d)$ and $\mathsf{allen}(\mathsf{dyspnea}, \mathsf{exacerbation}, [\mathcal{C} \cup \{\bar{o}\}])$ leading to the relation $\mathsf{allen}(\mathsf{exacerbation}, \mathsf{cough}, [\mathcal{C} \cup \mathcal{O}])$, with the distribution:

$$\begin{array}{lllll} o\ 0.009 & s\ 0.006 & d\ 0.006 & f\ 0.006 & eq\ 0.006 \\ \bar{o}\ 0.189 & \bar{s}\ 0.189 & \bar{d}\ 0.572 & \bar{f}\ 0.009 & \end{array} \tag{7.11}$$

resulting from the procedure described in 7.5.3 and using the uniform distribution to split up probabilities when the composition resulted in multiple basic relations. To compute $P(\mathsf{exacerbation})$ the relation is marginalised, but it nonetheless provides the insight that the probability that an exacerbation contains an episode of coughing is high.

Computing $P(\mathsf{exacerbation})$ from the rules and distributions above, we obtain $P(\mathsf{exacerbation}) = .001$, which is dominated by the small prior probability of an infection. Then, if we have evidence (from the patient) that dyspnea and cough are present, we obtain $P(\mathsf{exacerbation}) = 0.84$. See also Figure 7.5.

Dynamic Bayesian network

To show the different aspects of the language we can compare the situation above to a model based on a dynamic Bayesian network. As we have seen we can convert a model based on Allen's algebra to a DBN using the procedure in Section 7.5.3.

First we notice that possible infections are the repeating events (with relation *before*) that define the time slices of the DBN. Inspecting the rules above we see that the symptoms are contemporary with the infection event, resulting in connections within the time slice. Both observations are a direct consequence of Proposition 7.2. However, we also have that the relation between *dyspnea* and *exacerbation* is *overlaps* with probability 0.2. In Proposition 7.3 we gave a characterisation of situations for which we can construct a DBN. In the current situation it means that the exacerbation and its symptoms would have to be modelled as interacting in the same time slice. This seems a reasonable solution, resulting in the DBN structure shown in Figure 7.6. Note that a consequence of this choice is that we lose the possibility to make a distinction between the start of the symptoms (dyspnea, cough) and the exacerbation. This shows that although a DBN can be constructed for this situation, the representation with Allen's algebra is more expressive.

With respect to the probabilistic parameters we can reuse the parameters from the logic representation. We did not display the rules that govern the probability of symptoms without an infection, which are needed to complete a conditional probability table, but they have a similar pattern as the rules shown. The uncertainty on the relations is however less easy to represent: since we aggregate all concurrent relations, an obvious solution would be to marginalise out the relations. Taking as

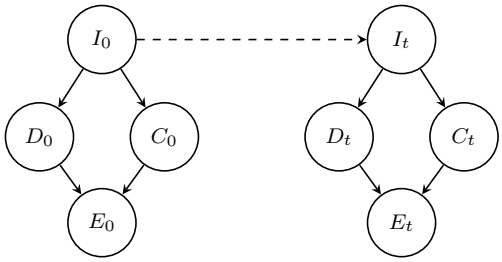


Figure 7.6: Structure of the COPD-monitoring DBN. Dashed arcs represent inter time-slice connections. I=infection, C=cough, D=dyspnea, E=exacerbation.

an example the probability of *exacerbation* given that *dyspnea* and *cough* are true, this results in the parameter estimate (omitting the conditioning on D, C on the right for readability):

$$\begin{aligned}
P(E \mid D, C) &= \sum_{R_{ED}, R_{EC}} P(E, R_{ED}, R_{EC}) \\
&= \sum_{R_{ED}, R_{EC}} P(E, R_{ED} \mid R_{EC}) P(R_{EC}) \\
&= \sum_{R_{ED} = \{\bar{o}, f\}} P(E, R_{ED}) + \sum_{R_{ED} = \{\mathcal{C} \setminus f\}} P(E, R_{ED}) \\
&= \sum_{R_{ED} = \{\bar{o}, f\}} P(E \mid R_{ED}) P(R_{ED}) + \sum_{R_{ED} = \{\mathcal{C} \setminus f\}} P(E \mid R_{ED}) P(R_{ED}) \\
&= 0.95 \cdot 0.55 + 0.7 \cdot 0.45
\end{aligned}$$

The parameter estimates for the other cases are similar.

Alternatively we can explicitly model the relations within the time slice by introducing extra variables that take the relations as values. To maintain the temporal semantics of the DBN we again only consider relations in $\mathcal{C}$ within a time slice. The network from Figure 7.6 can then be extended to the one in Figure 7.7. Since we assumed earlier that the distribution over temporal relations is independent of the values of the (event) variables involved, the relation variables do not have parents in the graph. The parameters of the relation variables are equal to the distribution in (7.4). For *exacerbation*, in this particular example R_{EC} has no influence and we can encode the joint distribution of *dyspnea*, *cough* and R_{ED} easily in the CPT. We obtain $P(E \mid D = 1, C = 1, R_{ED} = r) = p$ where $p = 0.95$ when $r \in \{\bar{o}, f\}$ and $p = 0.7$ when $r \in \{\mathcal{C} \setminus f\}$. This follows directly from the rules we gave earlier.

Summarising, we can see that comparing the two representations the DBN has a straightforward interpretation in terms of events that happen at the same time and those that do not. Allen's algebra allows a richer temporal structure without enforcing too much temporal detail. The logic representation can be used when only ordering information is available and allows us to model both structural and temporal information. Furthermore, since it is easy to define new relations for a particular domain situation, much more complex relations can also be specified easily. In order to gain the same level of detail in a DBN representation one has to introduce additional variables, losing some of the convenience of the compact representation of DBNs.

The best choice of temporal model is a balancing act where it depends on the importance and availability of temporal information whether a DBN or a logic rep-

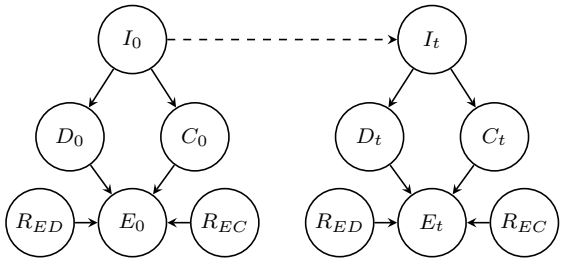


Figure 7.7: Structure of the COPD-monitoring DBN with variables for the temporal relations. Dashed arcs represent inter time-slice connections. I=infection, C=cough, D=dyspnea, E=exacerbation.

resentation has more value. In the case of COPD monitoring, a DBN is useful as a simple classifier that takes some temporal information into account. However, to gain a better understanding of the nature of the disease process it is worthwhile to use a richer representation that models different kinds of temporal relations between symptoms and between a symptom and an exacerbation. Using qualitative relations we can model the additional temporal structure, without requiring unrealistic temporal detail.

7.7 CONCLUSION

In this paper we reformulated and extended Allen's algebra in a probabilistic logic, leading to a framework called *Qualitative Time CP-logic*. Medical processes are often well suited to be represented in a qualitative time framework, and the representation explored in this paper appears to capture knowledge that fits clinical reasoning well. Specifically, the framework allows us to construct models stepwise, starting with structural properties represented by logic, adding qualitative temporal information and finally representing uncertainty with probabilities. Compared to directly constructing a dynamic Bayesian network, we gain the expressiveness of probabilistic logics in general and a more intuitive time representation. If desired, the model can still be converted to a DBN, but this requires some assumptions on how to simplify the temporal information represented with Allen's algebra. Our COPD model indicates that we can specify clinically relevant processes in an intuitive fashion within our framework, while the generic nature of the framework implies that we can also deal with many other biomedical situations.

8

DISCUSSION AND CONCLUSION

In this thesis we have studied both practical and theoretical aspects of dealing with uncertain temporal processes and data in a clinical context. We focussed on chronic disease management in which the combination of uncertainty and time occurs naturally, although similar situations can also be identified in other fields of medicine. The application domain of chronic disease care lead to the development of a monitoring system for COPD-exacerbation detection. The system has the potential to improve patient well-being and reduce health-care costs while at the same time provided a relevant context for the development and application of modelling techniques.

In this final chapter we place our contributions in a broader context and provide some views on directions for future research.

8.1 SMART MOBILE HEALTH

As was mentioned in the introduction, eHealth and more recently mHealth are a result of technological advances that allow us to automate some aspects of health care. Making information available is one of the main contributions of this automation. Our system also contributes to this aspect by making it possible for both the patient and the physician to be kept apprised of a patient's health status. However, as a result of monitoring too much information may become available, so that filtering becomes a necessity. We think that this is where our *smart* mobile health approach is an improvement over previous monitoring systems. Producing data is in itself not a useful activity unless it can be used to improve health care, either through better patient care or efficiency gains. For the former, it is important to give patients feedback on their health in a timely manner, for which our system is well suited. In addition feedback needs to be understandable and appropriate, a topic that in this stage of the development of our system has received fairly limited attention. To give proper feedback further cooperation with physicians is necessary to establish the appropriate course of action in a given situation, which is nowadays facilitated by clinical guidelines that summarise current best practices. With respect to efficiency we think that because early detection of exacerbations may prevent hospital admission and our system is better equipped to make timely detection possible than the methods that are now usual care, the potential for cost reduction is considerable.

Design

The design of the monitoring architecture was an important aspect to consider. While at the start of the project in 2008 it seemed reasonable to base the system on a notebook computer it quickly became apparent that a more future-proof solution would be implementing the system on a smartphone. The release of the Android operating system – or to be precise the availability of phones running Android in 2008 – provided the open-source platform that allowed us to easily develop our own

application. An advantage of using smartphones is the increased mobility and the fact that many people are already accustomed to carrying a phone on their person. A consequence of this choice was that we had to find sensors that could communicate with our application, which turned out to be difficult to find as medical sensors are usually only available as a monolithic package of hardware and software. Although we finally found a solution for spirometry and pulse-oximetry with Bluetooth communication, it is currently a rather custom and costly option. As an alternative we have also looked into developing our own hardware (in cooperation with the Radboud University Techno Centre). This has lead to a prototype of a new sensor interface, but more work is required before this can be used in practice. To enable large scale employment it is clear that both the hardware and software should be improved and tested further to ensure stable operation.

Model development

To implement smart mobile health care we have turned to probabilistic models to represent uncertain disease processes. The advantages of such models are that they can describe patient monitoring in a way that is easily interpretable; can be constructed both manually and from data; can be given a causal interpretation; are robust to missing values and are versatile in the sense that they can be used as a classifier as well as to provide insight in the domain by computing arbitrary probabilities of interest in the domain. These are all desirable properties which support the choice for Bayesian networks. This does not mean however, that a Bayesian network is the only possible solution, and indeed also for the COPD-exacerbation prediction task it is conceivable that other methods would be effective. Especially if we drop the requirement of interpretability, many statistical machine learning methods (see e.g. [46]) could be applied to classify monitoring data as either an exacerbation or not an exacerbation. With increasing model complexity however, including environment factors and therapy plans such that we can support patient self-management with mobile decision support, interpretability is of great importance to maintain accountability of the system. Bayesian networks are therefore in our opinion a good choice for automatic data interpretation.

The development of a disease model can be based on expert knowledge, on data or a combination of the two. Unfortunately, even though COPD is a common disease there is little data available. Clearly, particular data is available in patient records, but detailed day by day information is scarce and generally not accessible. Although patient data is privacy sensitive, one could argue here for open access of anonymised data for scientific purposes. The scarcity of data meant that hand-crafting a model was our only initial option, which does have the advantage of obtaining insightful knowledge from experts in the field. Knowledge elicitation is however a time-consuming and difficult task often hard to fit into the busy schedule of clinical specialists.

After the pilot study we had a limited amount of data, gathered with our system, which could then be exploited to learn models, leading to the work in Chapter 5 on model learning from small data samples. A possible shortcoming of structure learning temporal Bayesian networks from small data samples is that we may learn models that have too much variance. Although the first validation indicates that performance is quite good, a more extensive evaluation is necessary. Further limitations lie in the variables that are taken into account to predict exacerbations. To increase the practical clinical relevance the model needs to be extended from purely interpreting the monitoring data to taking into account patient history, therapy plans and environmental factors. For decision support purposes, especially therapy plans will be an important addition to the model.

Evaluation

The system as a whole and the model in particular needs to be evaluated to establish first whether it works as expected and second whether it has added value in health care. The pilot study served to start answering the first question, indicating that monitoring can be effective and that the model is capable of predicting exacerbations. The performance of the prediction degrades when we try to make predictions further ahead in time, which is to be expected, but gives us a kind of graceful degradation. For sudden onset exacerbations this is not the case however, as these cannot be detected before the actual onset. Although promising, the pilot study is not sufficient to fully establish these findings due to the small number of enrolled patients, so the generalisability is limited. Further clinical validation of the probabilistic model is necessary to make the step to actual decision support based on the predictions. The prototype does inspire confidence in the underlying ideas and consolidating the system to be practically clinically usable appears worthwhile. In the next section, on future research, we will outline further evaluation steps.

Further considerations

The nature of monitoring also has an ethical aspect as it encroaches on patients' private life. Ethics play a role in all medical procedures, but gathering data over a prolonged period of time at home is a particular kind of situation. Data security is an important issue that should be considered and furthermore, besides the purely technical perspective, it should be regulated who owns the data and who has access. We would argue here that patient empowerment should also mean that the patient is in control of the data and has to actively grant access rights (also for anonymised data for scientific purposes). Nonetheless from a clinical efficacy point of view it appears useful to integrate our system with existing health care information systems used by physicians and nurses. This is the usual trade off between security and convenience. In practice a middle ground will have to be found.

Embedding smart monitoring in actual clinical practice is not trivial. The road from a working prototype to a system that is implemented within existing health care procedures is full of pitfalls of a diverse nature related to responsibility, (technical) support, regulatory requirements, financing, etcetera. We should therefore be careful to keep the needs of the patient in mind and try to minimise procedural overhead. Focussing on supporting self-management can play a role in ensuring we really improve disease management from a patient perspective, as it is in principle possible for our system to function independently. Nonetheless integration with existing health care structures may have added value by improving information provision. This means an appropriate way has to be found to integrate our system, which in some sense is yet another information system, within the workflow of clinicians. Finally, for a practically useful system both patients and physicians need to be convinced of its use and usability, but we think that the potential benefits are sufficient to attempt to overcome the problems of implementation.

8.2 FUTURE RESEARCH

The work described in this thesis leaves various opportunities for further research. From a clinical perspective it is clear that the current system is promising but lacks a rigorous validation in the form of a randomised clinical trial designed to evaluate the benefits of the home monitoring system compared to usual care. At the time of writing, realising this validation is in progress, starting with a data gathering study with a larger group of patients to validate the probabilistic model. The availability

of additional data will also give us the opportunity to further investigate the prediction task and revisit the machine learning approach to find models and estimate parameters. One of the interesting questions that we will try to answer is whether longer term temporal trends can be found that improve prediction performance.

A possible improvement to the models used to capture the COPD monitoring process is including more general risk factors in the model to augment the pure monitoring data. Especially modelling in some detail the different therapies and their effect is of interest to capture the possible influence on future exacerbations. This has the added benefit of allowing the model to generate *what-if* scenarios for therapy change to compare their effect according to the model, which is useful for decision support. Clearly it is not feasible to model in detail the effect of the many drugs in use, but it does appear useful to model effects at the level of classes of drugs like short or long acting bronchodilators, corticosteroids and antibiotics.

The extension to decision support requires integration of patient advice in the model. Although related, this does not necessarily depend on modelling therapy effects, as we could also view therapy as a black box that relies on clinical guidelines and input from a physician. Various techniques have been developed over the years for decision making under uncertainty, notably Markov decision processes (MDPs) and influence diagrams, where the former has a focus on finding optimal policies for making many decisions while the latter usually deals with a more structured domain and can be viewed as an extension of Bayesian networks with decisions. Research in the integration of MDPs and dynamic Bayesian networks [13] is of interest as a way to extend our dynamic Bayesian network models to include decision making under uncertainty.

The characteristics of monitoring data described in Section 1.4 inspire further ideas for future work. We have studied some aspects of temporal indeterminacy in Chapter 6, but the problem is not yet solved. Studying more general ways to deal with indeterminacy, without relying on predetermined granularity bounds is still a challenge in terms of modelling, inference and learning. A generalisation to multiple granularities would also be of interest. The language we developed in Chapter 7 is a different approach to modelling temporal uncertainty but is limited to reasoning about the relations between events and not the actual interval instances, in the sense that it remains a qualitative representation. In many situations this turns out to be sufficient, but it may be possible to further extend the language. Irregular data acquisition [80] and non-stationary models [81] have received some attention but a general approach to tackle the combination of problems associated with representing uncertain temporal data is still of interest. It appears that probabilistic models or perhaps a generalisation to probabilistic logic are a useful starting point for this research.

Another topic that justifies further research is personalisation. To properly support each individual patient the system should be able to adapt to the patient's personal characteristics. Part of the value lies in the departure from group level models that are common in medicine. Especially for COPD that exhibits different disease types (number of exacerbations, bronchitic or emphysemic characteristics), personalisation may be able to improve decision support. Practically, personalisation may be achieved by using the monitoring data to update the parameters of the model. What the best way is to do so, taking into account that exacerbations are rare events, is an open research question.

Finally, a note on generalisability. We have developed a monitoring platform consisting of a smartphone application that includes a disease model, sensors and a web-application for administration purposes. To use this for a monitoring system targeted at a different disease some parts can be reused as the general architecture will be the same. However if different sensors are required new interfaces will have to be developed. The most important aspect for decision support is the disease model,

which is also the main bottleneck in terms of development. It usually requires a substantial investment of time from a domain expert and availability of data. So although the principles behind the system are generalisable, expansion to a different domain still entails a considerable amount of work. Nevertheless, mHealth solutions with disease models remain an interesting direction for research with great potential to improve healthcare.

BIBLIOGRAPHY

[1] S. Aaron, G. Donaldson, G. Whitmore, J. Hurst, T. Ramsay, and J. Wedzicha. Time course and pattern of COPD exacerbation onset. *Thorax*, 67(3):238–243, 2012.

[2] S. Acid, L. de Campos, J. Fernández-Luna, S. Rodríguez, J. Rodríguez, and J. Salcedo. A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine*, 30(3):215–232, 2004.

[3] J. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[4] J. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154, 1984.

[5] N. R. Anthonisen, J. Manfreda, C. P. W. Warren, E. S. Hershfield, G. K. M. Harding, and N. A. Nelson. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Annals of internal medicine*, 106(2):196–204, 1987.

[6] G. Arroyo-Figueroa and L. Sucar. A temporal Bayesian network for diagnosis and prediction. In K. Laskey and H. Prade, editors, *UAI '99: Proceedings of the 15th conference on Uncertainty in Artificial Intelligence*, pages 13–20. Morgan Kaufmann, San Francisco, CA, USA, 1999.

[7] G. Arroyo-Figueroa and L. Sucar. Temporal bayesian networks of events for diagnosis and prediction in dynamic domains. *Applied Intelligence*, 23(2):77–86, 2005.

[8] F. Bacchus, J. Tenenberg, and J. Koomen. A non-reified temporal logic. *Artificial Intelligence*, 52(1):87–108, 1991.

[9] A. Bangor, P. Kortum, and J. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

[10] J. Basilakis, N. Lovell, S. Redmond, and B. Celler. Design of a decision-support architecture for management of remotely monitored patients. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1216–1226, 2010.

[11] C. Bettini, X. Wang, and S. Jajodia. A general framework for time granularity and its application to temporal reasoning. *Annals of Mathematics and Artificial Intelligence*, 22:29–58, 1998.

[12] E. Bischoff, L. Boer, J. Molema, R. Akkermans, C. van Weel, J. Vercoulen, and T. Schermer. Validity of an automated telephonic system to assess COPD exacerbation rates. *Eur Respir J*, 39(5):1090–1096, 2012.

[13] C. Boutilier, R. Dearden, and M. Goldszmidt. Stochastic dynamic programming with factored representations. *Artificial Intelligence*, 121(1-2):49–107, 2000.

[14] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[15] G. Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[16] J. Brooke. SUS: A quick and dirty usability scale. In P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, editors, *Usability evaluation in industry*, 1996.

[17] A. Canty and B. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2012. `http://cran.r-project.org/web/packages/boot/`.

[18] P. Cheeseman. In defense of probability. In *IJCAI '85: Proceedings of the 9th International Joint Conference on Artificial Intelligence*, 1985.

[19] C. Combi, M. Franceschet, and A. Peron. Representing and reasoning about temporal granularities. *Journal of Logic and Computation*, 14(1):51–77, 2004.

[20] C. Combi, E. Keravnou, and Y. Shahar. *Temporal Information Systems in Medicine*. Springer Publishing Company Incorporated, New York, NY, USA, 2010.

[21] C. Combi, F. Pinciroli, and G. Pozzi. Managing different time granularities of clinical information by an interval-based temporal data model. *Methods of Information in Medicine*, 34(5):458–474, 1995.

[22] C. Combi and G. Pozzi. HMAP – A temporal data model managing intervals with different granularities and indeterminacy from natural language sentences. *The VLDB Journal*, 9(4):294–311, 2001.

[23] G. Cooper. A diagnostic method that uses causal knowledge and linear programming in the application of Bayes' formula. *Computer Methods and Programs in Biomedicine*, 22(2):223–237, 1986.

[24] F. Cozman. *EBayes: Embedded Bayesian Networks*. `http://www.cs.cmu.edu/~javabayes/EBayes/index.html/`.

[25] P. Dagum and A. Galper. Forecasting sleep apnea with dynamic models. In *UAI '93: Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 64–71, 1993.

[26] P. Dagum, A. Galper, and E. Horvitz. Dynamic network models for forecasting. In *UAI '92: Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, pages 41–48, 1992.

[27] D. Dash and M. Druzdzel. A hybrid anytime algorithm for the construction of causal models from sparse data. In *UAI '99: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 1999.

[28] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3):142–150, 1989.

[29] R. Dechter, I. Meiri, and J. Pearl. Temporal constraint networks. *Artificial Intelligence*, 49(1-3):61–95, 1991.

[30] Decision Systems Laboratory, University of Pittsburgh. *SMILE: Structural Modeling, Inference, and Learning Engine*. `http://genie.sis.pitt.edu/`.

[31] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[32] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[33] G. Elidan and N. Friedman. Learning the dimensionality of hidden variables. In *UAI '01: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 2001.

[34] S. Evers, M. Fokkinga, and P. Apers. Probabilistic processing of interval-valued sensor data. In *Proceedings of the 5th workshop on Data management for sensor networks*, 2008.

[35] N. Friedman. The Bayesian structural EM algorithm. In *UAI '98: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998.

[36] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[37] N. Friedman, M. Goldszmidt, and A. Wyner. Data analysis with Bayesian networks: a bootstrap approach. In *UAI '99: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 1999.

[38] N. Friedman, K. Murphy, and S. Russell. Learning the structure of dynamic probabilistic networks. In *UAI '98: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998.

[39] S. Galán, F. Aguado, F. Díez, and J. Mira. NasoNet, modelling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. *Artificial Intelligence in Medicine*, 25(3):247–264, 2002.

[40] GOLD. Global Initiative for Obstructive Lung Disease. `www.goldcopd.com`.

[41] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

[42] M. Grzegorczyk and D. Husmeier. Non-homogeneous dynamic Bayesian networks for continuous data. *Machine Learning*, 83(3):355–419, 2011.

[43] B. Gutmann, I. Thon, A. Kimmig, M. Bruynooghe, and L. de Raedt. The magic of logical inference in probabilistic programming. *Theory and Practice of Logic Programming*, 11:663–680, 2011.

[44] D. Halpin, T. Laing-Morton, S. Spedding, M. Levy, P. Coyle, J. Lewis, P. Newbold, and P. Marno. A randomised controlled trial of the effect of automated interactive calling combined with a health risk forecast on frequency and severity of exacerbations of COPD assessed clinically and using EXACT PRO. *Prim Care Respir J*, 20(3):324–331, 2011.

[45] A. Hartemink. *Banjo: Bayesian Network Inference with Java Objects*. Duke University, 2010. `http://www.cs.duke.edu/~amink/software/banjo/`.

[46] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer-Verlag, 2008.

[47] D. Heckerman, D. Geiger, and D. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.

[48] P. Hernandez-Leal, A. Rios-Flores, S. Ávila-Rios, G. Reyes-Terán, J. González, F. Orihuela-Espina, E. Morales, and L. Sucar. Discovering human immunodeficiency virus mutational pathways using temporal Bayesian networks. *Artificial Intelligence in Medicine*, 57(3):185–195, 2013.

[49] B. Himes, Y. Dai, I. Kohane, S. Weiss, and M. Ramoni. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *Journal of the American Medical Informatics Association*, 16(3):371–379, 2009.

[50] A. Hommersom and P. Lucas. Generalising the interaction rules in probabilistic logic. In T. Walsh, editor, *IJCAI '11: Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 912–917. AAAI Press, 2011.

[51] J. Hurst, G. Donaldson, J. Quint, J. Goldring, A. Patel, and J. Wedzicha. Domiciliary pulse-oximetry at exacerbation of chronic obstructive pulmonary disease: prospective pilot study. *BMC Pulm Med*, 10(1):52, 2010.

[52] P. Jonsson and C. Bäckström. A unifying approach to temporal constraint reasoning. *Artificial Intelligence*, 102(1):143–155, 1998.

[53] H. Kautz and P. Ladkin. Integrating metric and qualitative temporal reasoning. In T. Dean and K. McKeown, editors, *AAAI '91: Proceedings of the 9th National Conference of the American Association for Artificial Intelligence*, pages 241–246. AAAI Press, 1991.

[54] M. Kearns, Y. Mansour, and A. Ng. An information-theoretic analysis of hard and soft assignment methods for clustering. In *UAI '97: Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, 1997.

[55] E. Keravnou. A multidimensional and multigranular model of timefor medical knowledge-based systems. *Journal of Intelligent Information Systems*, 13:73–120, 1999.

[56] S. Kim, S. Imoto, and S. Miyano. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Briefings in Bioinformatics*, 4(3):228–235, 2003.

[57] D. Koller and N. Friedman. *Probabilistic graphical models: Principles and techniques*. The MIT Press, Cambridge, MA, USA, 2009.

[58] A. Krokhin, P. Jeavons, and P. Jonsson. Reasoning about temporal relations: The tractable subalgebras of Allen's interval algebra. *Journal of the ACM*, 50(5):591–640, 2003.

[59] H. Lähdesmäki and I. Shmulevich. Learning the structure of dynamic Bayesian networks from time series and steady state measurements. *Machine Learning*, 71(2-3):185–217, 2008.

[60] S. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999.

[61] B. Lijnse and R. Plasmeijer. iTasks 2: iTasks for end-users. In *Proceedings of the 21st Symposium on Implementation and Application of Functional Languages*, 2009.

[62] P. Lucas, H. Boot, and B. Taal. Computer-based decision-support in the management of primary gastric non-hodgkin lymphoma. *Methods of Information in Medicine*, 37:206–219, 1998.

[63] P. Lucas, L. van der Gaag, and A. Abu-Hanna. Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine*, 30:201–214, 2004.

[64] C. Maiolo, E. Mohamed, C. Fiorani, and A. D. Lorenzo. Home telemonitoring for patients with severe respiratory illness: the Italian experience. *J Telemed Telecare*, 9(2):67–71, 2003.

[65] D. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6(2):101–155, 1982.

[66] B. McKinstry, H. Pinnock, and A. Sheikh. Telemedicine for management of patients with COPD? *The Lancet*, 374(9691):672–673, 2009.

[67] S. McLean, U. Nurmatov, J. Liu, C. Pagliari, J. Car, and A. Sheikh. Telehealthcare for chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews*, 7, 2011.

[68] K. Murphy. *Machine Learning: a Probabilistic Perspective*. The MIT Press, Cambridge, MA, USA, 2012.

[69] K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, USA, 2002.

[70] B. Nebel and H.-J. Bürckert. Reasoning about temporal relations: A maximal tractable subclass of Allen's interval algebra. *Journal of the ACM*, 42(1):43–66, 1995.

[71] D. Niewoehner. The impact of severe exacerbations on quality of life and the clinical course of chronic obstructive pulmonary disease. *The American Journal of Medicine*, 119(10, S1):38 – 45, 2006.

[72] U. Nodelman, C. Shelton, and D. Koller. Continuous-time Bayesian networks. In *UAI '02: Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002.

[73] H. Oh, C. Rizo, M. Enkin, and A. Jadad. What is ehealth?: a systematic review of published definitions. *J Med Internet Res*, 7(1):e1, 2005.

[74] G. Paré, C. Sicotte, D. St Jules, and R. Gauthier. Cost-minimization analysis of a telehomecare program for patients with chronic obstructive pulmonary disease. *Telemed J e-Health*, 12(2):114–121, 2006.

[75] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[76] J. Pearl. *Causality - Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.

[77] L. Peelen, N. de Keizer, E. de Jonge, R. Bosman, A. Abu-Hanna, and N. Peek. Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit. *Journal of Biomedical Informatics*, 43(2):273–286, 2010.

[78] S. Peirce, A. Hardisty, A. Preece, and G. Elwyn. Designing and implementing telemonitoring for early detection of deterioration in chronic disease: defining the requirements. *Health Informatics Journal*, 17(3):173–190, 2011.

[79] J. Polisena, K. Tran, K. Cimon, B. Hutton, S. McGill, K. Palmer, and R. Scott. Home telehealth for chronic obstrcutive pulmonary disease: a systematic review and meta-analysis. *J Telemed Telecare*, 16(3):120–127, 2010.

[80] M. Ramati and Y. Shahar. Irregular-time Bayesian networks. In *UAI '10: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 2010.

[81] J. Robinson and A. Hartemink. Learning non-stationary dynamic Bayesian networks. *Journal of Machine Learning Research*, 11:3647–3680, 2010.

[82] V. Ryabov and A. Trudel. Probabilistic temporal interval networks. In C. Combi and G. Ligozat, editors, *TIME '04: Proceedings of the 11th International Symposium on Temporal Representation and Reasoning*, pages 64–67, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[83] E. Santos Jr. and J. Young. Probabilistic temporal networks: A unified framework for reasoning with time and uncertainty. *International Journal of Approximate Reasoning*, 20(3):263–291, 1999.

[84] T. Sato. A statistical learning method for logic programs with distribution semantics. In L. Sterling, editor, *ICLP '95: Proceedings of the 12th International Conference on Logic Programming*, pages 715–729, Cambridge, MA, USA, 1995. The MIT Press.

[85] S. Selvin. *Survival Analysis for Epidemiologic and Medical Research*. Cambridge University Press, 2008.

[86] G. Shafer. *The Art of Causal Conjecture*. The MIT Press, Cambridge, MA, USA, 1996.

[87] Y. Shahar. A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90(1-2):79–133, 1997.

[88] E. Shortliffe and B. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23(3-4):351–379, 1975.

[89] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. The MIT Press, Cambridge, MA, USA, 2000.

[90] S. Srinivas. A generalization of the noisy-or model. In D. Heckerman and E. Mamdani, editors, *UAI '93: Proceedings of the 9th conference on Uncertainty in Artificial Intelligence*, pages 208–218. Morgan Kaufmann, San Francisco, CA, USA, 1993.

[91] A. Tawfik and E. Neufeld. Temporal reasoning and Bayesian networks. *Computational Intelligence*, 16(3):349–377, 2000.

[92] P. Terenziani, C. Carlini, and S. Montani. Towards a comprehensive treatment of temporal constraints in clinical guidelines. In A. Artale and M. Fisher, editors, *TIME '02: Proceedings of the 9th International Symposium on Temporal Representation and Reasoning*, pages 20–27, Los Alamitos, CA, USA, 2002. IEEE Computer Society.

[93] I. Thon, N. Landwehr, and L. de Raedt. Stochastic relational processes: Efficient inference and applications. *Machine Learning*, 82(2):239–272, 2011.

[94] J. Trappenburg, A. Niesink, G. de Weert-van Oene, H. van der Zeijden, R. van Snippenburg, A. Peters, J. Lammers, and A. Schrijvers. Effects of telemonitoring in patients with chronic obstructive pulmonary disease. *Telemed J e-Health*, 14(2):138–146, 2008.

[95] A. Tucker and X. Liu. Learning dynamic Bayesian networks from multivariate time series with changing dependencies. In *IDA '03: Proceedings of the 5th International Symposium on Intelligent Data Analysis*, volume 2810 of *LNCS*, pages 100–110, 2003.

[96] M. van der Heijden, B. Lijnse, P. Lucas, Y. Heijdra, and T. Schermer. Managing COPD exacerbations with telemedicine. In M. Peleg, N. Lavrač, and C. Combi, editors, *AIME '11: Proceedings of the 13th Conference on Artificial Intelligence in Medicine*, volume 6747 of *LNAI*, pages 169–178, 2011.

[97] M. van der Heijden, P. Lucas, B. Lijnse, Y. Heijdra, and T. Schermer. An autonomous mobile system for the management of COPD. *Journal of Biomedical Informatics*, 46(3):458–469, 2013.

[98] M. van der Heijden and P. J. Lucas. Describing disease processes using a probabilistic logic of qualitative time. *Artificial Intelligence in Medicine*, in press.

[99] M. van der Heijden and P. J. Lucas. Probabilistic reasoning with temporal indeterminacy. In *PGM '12: Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, 2012.

[100] M. van der Heijden, P. J. Lucas, and T. R. Schermer. Automatic COPD-exacerbation prediction: findings from a home-monitoring pilot. *Respiratory Medicine*, submitted.

[101] M. van der Heijden, M. Velikova, and P. J. Lucas. Learning Bayesian networks for clinical time series analysis. *Journal of Biomedical Informatics*, submitted.

[102] M. van Gerven, B. Taal, and P. Lucas. Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics*, 41(4):515–529, 2008.

[103] J. Vennekens, M. Denecker, and M. Bruynooghe. CP-logic: A language of causal probabilistic events and its relation to logic programming. *Theory and Practice of Logic Programming*, 9(3):245–308, 2009.

[104] G. Viegi, F. Pistelli, D. Sherrill, S. Maio, S. Baldacci, and L. Carrozzi. Definition, epidemiology and natural history of COPD. *Eur Respir J*, 30(5):993–1013, 2007.

[105] L. Vila. A survey on temporal reasoning in artificial intelligence. *AI Communications*, 7(1):4–28, 1994.

[106] M. Vitacca, G. Assoni, P. Pizzocaro, A. Guerra, L. Marchina, S. Scalvini, F. Glisenti, A. Spanevello, L. Bianchi, L. Barbano, A. Giordano, and B. Balbi. A pilot study of nurse-led, home monitoring for patients with chronic respiratory failure and with mechanical ventilation assistance. *J Telemed Telecare*, 12(7):337–342, 2006.

[107] Th. Vontetsianos, P. Giovas, Th. Katsaras, A. Rifopoulou, G. Mpirmpa, P. Giaboudakis, S. Koyrelea, G. Kontopyrgias, and B. Tsoulkas. Telemedicine-assisted home support for patients with advanced chronic obstructive pulmonary disease: preliminary results after nine-month follow-up. *J Telemed Telecare*, 11(S1):86–88, 2005.

[108] J. Wedzicha and T. Seemungal. COPD exacerbations: defining their cause and prevention. *The Lancet*, 370(9589):786–796, 2007.

[109] T. Wilkinson, G. Donaldson, J. Hurst, T. Seemungal, and J. Wedzicha. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 169(12):1298–1303, 2004.

[110] World Health Organisation. Chronic obstructive pulmonary disease factsheet. `http://www.who.int/mediacentre/factsheets/fs315/en/index.html`, 2012.

[111] A. Yañez, D. Guerrero, R. Pérez de Alejo, F. Garcia-Rio, J. Alvarez-Sala, M. Calle-Rubio, R. Malo de Molina, M. Falcones, P. Ussetti, J. Sauleda, E. García, J. Rodríguez-González-Moro, M. Gay, M. Torrent, and A. Agustí. Monitoring breathing rate at home allows early identification of COPD exacerbations. *CHEST Journal*, 142(6):1524–1529, 2012.

[112] M. Zweig and G. Campbell. Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, 39(4):561–577, 1993.

SUMMARY

Uncertainty is one of the difficulties one has to deal with in complex decision making situations. In this thesis we have studied this problem using probabilistic models in the clinical context of chronic disease management. Chronic obstructive pulmonary disease (COPD) is one of the major chronic diseases worldwide. COPD-patients suffer from a progressive lung disease with episodes of acute worsening of symptoms called *exacerbations*. Exacerbations have great impact on patient well-being and on health-care costs due to hospital admission. We have developed a system which aims to reduce exacerbation impact by monitoring patients at home and automatically interpreting the data to detect exacerbations. The system consists of a smartphone application that registers symptoms via a questionnaire and sensors to assess lung function. To predict the occurrence of exacerbations taking into account the uncertainty in the measurements, we used probabilistic models constructed in cooperation with pulmonologists and from patient data. We studied different versions of the model, using methods suitable for limited amounts of data, explicitly taking the time aspect of patient monitoring into account. A pilot study in which ten patients used our system suggests that we are able to predict the onset of an exacerbation a day in advance for gradual onset exacerbations and detect sudden onset exacerbations on the first day. After further testing our system could be used to support COPD patients by providing direct feedback based on the automatic interpretation of their monitoring data, which hopefully prevents hospital admission due to COPD. In this thesis we have further studied some theoretical properties of reasoning about uncertain time processes, of which patient monitoring is an example. We have developed a framework based on probabilistic logic that allows reasoning about time in a qualitative way using relations like 'hospital admission occurred during an exacerbation'. This kind of reasoning is fairly natural when precise timing information is unavailable but ordering is of interest. In this thesis we have made some progress both theoretically and practically in modelling uncertain temporal processes, however numerous options for future research in knowledge representation and clinical applications remain.

SAMENVATTING

Onzekerheid is een van de problemen waar we mee om moeten gaan wanneer we complexe beslissingen nemen. In dit proefschrift hebben we dit probleem onderzocht, met behulp van probabilistische modellen, in de klinische context van chronische ziekten. Chronisch obstructief longlijden (Engelse afkorting: COPD) is een van de meest voorkomende chronische ziekten ter wereld. COPD-patiënten lijden aan een voortschrijdende longziekte met episodes van acuut toenemende symptomen, *exacerbaties* genoemd. Exacerbaties hebben grote invloed op patiëntwelzijn en op zorgkosten veroorzaakt door ziekenhuisopname. Wij hebben een systeem ontwikkeld met als doel de impact van exacerbaties te verminderen door patiënten thuis te monitoren en automatisch de verzamelde data te interpreteren om exacerbaties te detecteren. Het systeem bestaat uit een smartphone-applicatie die symptomen registreert met behulp van een vragenlijst en sensoren om longfunctie te meten. Om exacerbaties te voorspellen, waarbij rekening wordt gehouden met de onzekerheid in de metingen, gebruiken we kansmodellen ontwikkeld in samenwerking met longartsen en gebruikmakend van patiëntdata. Verschillende versies van de modellen zijn bestudeerd met methoden geschikt voor het maken van modellen met weinig data, waarbij expliciet rekening wordt gehouden met het tijdsaspect van patiëntmonitoring. Een pilotstudie met tien patiënten suggereert dat we in staat zijn om de start van exacerbaties een dag van tevoren te voorspellen in het geval van geleidelijk ontwikkelende exacerbaties en te detecteren op de eerste dag in het geval van een plotselinge start. Na verdere evaluatie zou ons systeem gebruikt kunnen worden om COPD-patiënten te ondersteunen door directe terugkoppeling te geven gebaseerd op de automatische interpretatie van de monitoringdata om hopelijk daarmee ziekenhuisopname door COPD te voorkomen. In dit proefschrift bestudeerden we verder theoretische eigenschappen van het redeneren over onzekere tijdsprocesses, waarvan patiëntmonitoring een voorbeeld is. We hebben een raamwerk ontwikkeld gebaseerd op probabilistische logica waarmee we op een kwalitatieve manier kunnen redeneren over tijd met behulp van relaties zoals 'opname vond plaats gedurende een exacerbatie'. Deze vorm van redeneren is redelijk natuurlijk als precieze tijdsinformatie niet voorhanden is maar de ordening van gebeurtenissen van belang is. In dit proefschrift hebben we zowel theoretisch als praktisch enige progressie gemaakt in het modelleren van onzekere tijdsprocessen, maar er bestaan nog talrijke mogelijkheden voor toekomstig onderzoek naar kennisrepresentatie en klinische toepassingen.

DANKWOORD

Dit proefschrift zou niet tot stand zijn gekomen zonder bijdrage van allerlei mensen om mij heen. Ten eerste ben ik dank verschuldigd aan mijn promotor, Peter, je hebt me veel geleerd de afgelopen jaren, zonder jou zou dit proefschrift er niet geweest zijn. Je had veel goede ideeën en de inhoudelijke samenwerking heb ik erg kunnen waarderen. Tjard, jouw kennis over COPD en de medische praktijk waren van groot belang om dit interdisciplinaire project te voorzien van de nodige praktijkkennis. Dank voor het meedenken en de begeleiding.

Allerlei mensen hebben meegewerkt aan verschillende delen van het veelomvattende project rond COPD, waarvoor mijn dank. Met het risico mensen te vergeten, noem ik toch een paar namen: Johan Molema, Jan Vercoulen, Yvonne Heijdra, die specialistische kennis hebben bijgedragen aan het COPD-model; Riet Cretier die me enorm geholpen heeft met de praktische uitvoering van de pilot studie; de clinici en verpleegkundigen die hebben meegewerkt aan de pilot in Almelo, Arnhem en Groesbeek; en Bas Lijnse die zowel heeft meegedacht als uitgevoerd om het monitoring systeem te maken.

Graag bedank ik ook mijn kamergenoten door de jaren heen (ik herinner me leuke discussies op vrijdagmiddag), en andere collega's bij ICIS waar ik mee heb samengewerkt of anderszins mee te maken heb gehad. In alfabetische volgorde: Arjen, Bas, Eduard, Freek, Marina, Martijn, Max, Maya, Sander, Saskia, Steffen, Suzan, Theo, Thomas, Wenyun, Wessel: dank.

Bram en Lies, fijn dat jullie paranimfen wilden zijn. Merel, dankje voor de hulp met de vormgeving.

Dank ook aan de vele vrienden en kennissen voor het aanhoren van mijn misschien wat oninteressante verhalen over onderzoek en promotieperikelen en voor de af en toe noodzakelijke afleiding. Jullie zijn met te veel om op te noemen en ik zou alleen maar mensen vergeten, maar het werd (en wordt) gewaardeerd.

Rest mij slechts mijn ouders en broer te bedanken voor de getoonde interesse – ondanks dat het soms moeilijk uit te leggen was wat ik nu eigenlijk aan het doen was – de aanmoedigingen en ondersteuning: dankjewel!

CURRICULUM VITAE

Maarten van der Heijden

1984 born in Nijmegen.

1996-2002 Gymnasium, Dominicus College Nijmegen.

2002-2005 Bachelor Artificial Intelligence,
Radboud University Nijmegen.

2005-2007 Research Master Cognitive Neuroscience,
Radboud University Nijmegen.

2005-2008 Master Artificial Intelligence,
Radboud University Nijmegen.

2008-2013 PhD-candidate,
Department of Primary and Community Care, Radboud University Nijmegen Medical Centre;
Model-Based System Development, Institute for Computing and Information Sciences, Radboud University Nijmegen.

“Breath in a second, wait, wait, you’ll be okay.”